ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 14, No. 9, pp. 2709-2721, September 2024
DOI: https://doi.org/10.17507/tpls.1409.07

Exploring Lexical Luxury in Scott Fitzgerald's Novels: A Corpus-Based Stylistic Analysis

Hasan Ali Hussein
Department of English, College of Education for Human Science, University of Babylon, Babylon, Iraq;
Nasiriyah Technical Institute, Southern Technical University, Di-Qar, Iraq

Nesaem M. Al-Aadili
Department of English, College of Education for Human Science, University of Babylon, Babylon, Iraq

***Abstract*—This study is based on corpus-stylistic techniques and measures the concept of lexical luxury being considered as a stylistic marker in literary work represented by Scott Fitzgerald's novels. This study aims to investigate the feasibility of measuring lexical luxury using corpus-based methodologies and techniques. It looks at how lexical luxury is used as a stylistic marker in literary works. The study hypothesizes that corpus stylistics frames a rather reliable source of verifying the validity of controversial issues about lexical luxury measurements; Type/Token Ratio is a measure that can efficiently determine the size of the lexical luxury when an analyst analyzes a text; the high percentages of lexical luxury are based on the high concentrations of lexical items resulting in a denser lexicality. The corpus in this study is designed and compiled on the principles of corpus linguistics as well as corpus stylistics, then, processed via *WordjSmith Tools* (8.0) to get the *tokens*; *types*, and *word frequencies*. The lexical items and function words are analyzed using Ure's method, then percentages of lexical luxury are obtained. The percentages represent lexical luxury profiles for each novel. After that, the whole statistical output is graphically plotted using Microsoft Office Excel. This study found that corpus-based stylistic studies have proven to be effective in exploring specific aspects of style. However, corpus stylistics makes significant use of statistical descriptions offered by trustworthy and authentic measures such as Type/Token Ratio.**

***Index Terms*—lexical luxury, stylistic marker, corpus linguistics, corpus stylistics, Type/Token Ratio**

I. Introduction

Corpus linguistics provides a collection of statistical procedures to determine the number of items. Following that, the numbers must be put into a practical formula to measure the lexical luxury of a given text. The level of lexicality determined by the statistical formula could indicate that lexical luxury is a stylistic marker for determining the lexical regression or progression of a specific writer's style.

Consequently, the present study focuses on adopting corpus-based techniques to measure a feature of lexical diversity. This feature is concerned with the identification of lexical luxury and its graphical representation. It likewise looks at how this feature is used as a stylistic marker in literary works. Consequently, the problem addressed in this work can be stated in the following questions: What role does lexical luxury play as a stylistic marker? What are the indices of measuring lexical luxury in texts? What are the attainable statistical frameworks proposed by corpus?

The present study aims to investigate the usefulness of the Type/Token Ratio (TTR) in measuring lexical luxury in five English novels written by Scott Fitzgerald, and finding their statistical reliability in giving rigorous accounts of lexical luxury measurements; comparing TTR values to figure out the size of the lexical luxury that occurred in the process of writing a text. Doing such a step, the researchers may hopefully give objective and scientific evidence of progression or recess in the linguistic competence of a writer; proving that lexical luxury is a stylistic marker. This study hypothesizes that corpus stylistics frames a rather reliable source of verifying the validity of controversial issues about lexical luxury measurements, and lexical luxury is a well-organized measure for measuring lexicality in literary texts and the lexical progression of a writer's style.

The present study is limited to investigating one specific type of lexical measure which is a lexical luxury, and it is limited to exploring one type of genre: novels represented by five English novels introduced by Scott Fitzgerald: *This Side of Paradise* (1920); *The Beautiful and Damned* (1922); *The Great Gatsby* (1925); *Tender Is the Night* (1934); and *The Last Tycoon* (1941).

II. Corpus Linguistics, Stylistics and Corpus Stylistics

The present section is an attempt to determine the relationship that regulates three fundamental areas of study: Corpus Linguistics, Stylistics, and Corpus Stylistics.

A. Corpus Linguistics

Corpus Linguistics (henceforth CL) is a field of study that focuses on procedures and methods for studying languages (McEnery & Hardie, 2012). CL is an empirical approach used to conduct investigations of language variability to arrive at scientifically valid conclusions. CL is an application methodology with a well-developed set of rules and theoretical foundations (Biber & Reppen, 2015).

Based on examples of language use in real-life situations, CL is the best way to study languages (McEnery & Wilson, 2001). CL can achieve its goals by using corpus data as a first step toward conducting a scientific and practical investigation (McEnery & Hardie, 2012). It involves a computer-based analysis of massive amounts of data, which means that CL is "multilingual", having various languages and dialects that may be investigated using corpus data (McEnery & Hardie, 2012). Thus, Hunston (2022) defines corpus as a collection of examples of language use that can be spoken or written, and CL is unquestionably a "store of used language" (p. 3).

B. Stylistics

Norgaard et al. (2010) state that stylistics explores how meaning is formed through language in literature and other types of writings or texts. Francis (2017) suggests that stylistics is developed by Russian Formalism via Prague Structuralism, which is followed by the concept of estrangement, or "deviation from normal usages" (p. 44). This shows that style is associated with a deviation from a norm, or the conventional usage of language, in order to achieve literary, persuasive, rhetorical, or other effects (Hickey, 1993). Thus, stylistics might be introduced as a subfield of linguistics that studies distinctive linguistic expressions or the style itself (Verdonk, 2002). It also tries to show critics and linguists the vast possibilities for interpretation (Simpson, 2004).

C. Corpus Stylistics

Corpus Stylistics (CS) is a stylistic offshoot in which corpus techniques are used to support the analysis of stylistic aesthetic function in large texts (McIntyre, 2012). CS is the most recent branch of stylistics that incorporates corpus linguistic tools into the stylistician's toolkit (Shepherd & Sardinha, 2013).

CS is now a major field of study in literary linguistics. It combines a "variety of methods, including the plain use of digitized literary texts, as well as the application of statistical analyses of literary works or contrastive and comparative corpora" (Wynne, 2006, p. 2). CS advancement coincides with the advancement of CL methodologies as they are used in the analysis of bulky amounts of data and literary texts. All of this is possible with computer software programs (Norgaard et al., 2010). Furthermore, CS can be viewed as a study of the language in literature. It is the application of models and theories from corpus linguistics and stylistics to machine-readable databases (McIntyre, 2015).

III. Lexical Luxury as a Stylistic Marker

Lexical luxury will be illustrated and disputed as a stylistic marker in this Section. This stylistic marker will be discussed in connection to the ideas of lexical frequency, lexical variety, and lexical richness.

A. The Concept of Stylistic Marker

Style in text analysis may be characterized as a "set of measurable patterns which may be unique to an author" (Holmes, 1994, p. 87). It is closely connected to choices that add diversity and depth to the text. The language features that make up those choices are known as stylistic markers. The stylistic markers are linguistic elements or features that are thought to be stylistically significant (Yumin, 1986).

When two texts are compared, one of these linguistic traits may occur in one text but not in the other. It may appear more frequently in one text than in the other, or it may occur at the same frequency in both texts (Enkvist, 1973). The frequencies or densities of these features deviate considerably from those in the norm (i.e., the set of expectations based on previous experiences) (Enkvist, 1973). Thus, stylistic markers are the distinctive features of a writer's style that are primarily utilized to identify stylistic features in a certain text (Yeibo, 2011).

In general, stylistic markers could be identified at any linguistic level. These markers can be found at the morphological, semantic, syntactic, and even text format levels (McMenamin, 2002). Furthermore, these markers cannot be examined with the naked eye (Eder, 2011). The only practical procedure for establishing the significance of such markers is the empirical investigation (Eder, 2011).

As a result, stylistic markers are a set of idiosyncrasies that arise in addition to uncommon linguistic features that may occur (McMenamin, 2002). Word frequencies, word length, sentence length, and so forth might be chosen and investigated as stylistic markers (Eder, 2011).

B. Lexical Luxury

Lexical Luxury (LL) denotes the content words in relation to the overall quantity of words in the text (Singla, 2012). LL refers to a text's "degree of richness in terms of ideas, meanings, and information" (Al-wahy, 2016, p. 5). Several writers demonstrate and discuss LL from various angles. Camiciottoli (2007) presents LL as the ratio of content words in relation to grammatical words. Likewise, Eggins (2004) defines lexical luxury in a sentence or a text by measuring the amount of content terms as a percentage of total words. In other words, LL is the proportion of lexical items to the overall running words in the text (Halliday, 1989).

As a result, LL may be measured by the proportion of lexical terms in a text (Keune, 2012). This percentage shows the frequency with which nouns, verbs, adverbs, and adjectives are used (Keune, 2012). Furthermore, Johansson (2008) notes that a text with a high lexical item ratio is more informative than one with a high function or grammatical word ratio. In this regard, LL is a marker of information weight in a certain text, "a text amount of details and technical vocabulary" (Giannossa, 2012, p. 48). Lastly, LL is calculated by dividing the total number of lexically featured words in a text or corpus by the total number of orthographic or grammatically featured words (Johansson, 2008).

IV. Corpus Design and Methodology

After having a thorough examination of lexical luxury as a stylistic marker, the methodology and techniques utilized in this study are now ready to be presented. This section highlights some crucial points about corpus design as well as the main processes that each researcher should follow to construct a satisfactory corpus. Each step is presented as a separate item hoping that readers will understand how a corpus can be constructed and what procedures should be followed when constructing a corpus.

A. Corpus Design

The first step in designing a corpus should be to plan. For Atkins et al. (1992), the design of a corpus must specify *the type* of corpus, *the size* of the corpus, *the period of time* involved, *the language varieties*, and *the mode* of data (written or spoken). Because the corpus is defined as a manifestation of a language, it must be built to meet the purpose it is meant to represent. Furthermore, for a more effective and well-documented corpus-based study, two components are critical: (1) software tools (for data reading); and (2) human intuition (for data interpretation) (Anthony, 2009).

(a). Corpus Data

The corpora (digital texts) have been collected from a website (www.getenberg.org) by downloading material from the World Wide Web. This website is the most trustworthy among others, and it is chosen for its well-known academic reputation. Throughout the study, one computerized corpus has been constructed; this corpus is comprised or compiled of English novels produced by the American novelist Scott Fitzgerald; they (the novels) are all in the same genre: fiction.

To obtain the raw material, certain types of data have been removed from the full texts, such as (websites, personal descriptions, numerals, and footnotes). The corpora are then divided into successive samples. The researchers employed an approximate number of words or tokens (i.e., roughly 5,000 tokens in size) to segment valid and representative text samples, using Biber's (1993) strategy. The data came up with a total of (83) samples after segmenting all of the corpora. As a result, the table below provides a statistical description of the corpora employed in this study.

TABLE 1
CORPORA DESCRIPTION OF THE SOURCE TEXTS TOKENS: *"THIS SIDE OF PARADISE"* (1920), *"THE BEAUTIFUL AND DAMNED"* (1922), *"THE GREAT GATSBY"* (1925), *"TENDER IS THE NIGHT"* (1934), AND *"THE LAST TYCOON"* (1941)

Samples No.	*This Side of Paradise*	*The Beautiful and Damned*	*The Great Gatsby*	*Tender Is The Night*	*The Last Tycoon*
1	5.006	5.003	5.026	5.003	5.001
2	5.006	5.002	5.020	5.001	5.006
3	5.008	5.005	5.002	5.002	5.000
4	5.002	5.000	5.006	5.006	5.002
5	5.006	5.002	5.008	5.002	5.005
6	5.000	5.007	5.005	5.018	5.008
7	5.004	5.001	5.000	5.017	5.004
8	5.003	5.009	5.011	5.028	5.007
9	5.002	5.006	5.001	5.020	5.003
10	5.005	5.003	2.803	5.006	1.966
11	5.007	5.008		5.047	
12	5.001	5.008		5.000	
13	5.001	5.005		5.012	
14	5.000	5.002		5.029	
15	5.006	5.003		5.009	
16	4.699	5.002		5.005	
17		5.007		5.007	
18		5.003		5.003	
19		5.008		5.008	
20		5.003		5.032	
21		5.004		5.011	
22		5.002		3.107	
23		5.004			
24		5.002			
25		3.380			
Average	79.756	123.479	47.882	108.373	47.002

(b). Selection Features

With the assistance of the World Wide Web, the full texts have been downloaded from the (www.getenberg.org) website. The downloaded texts for this study are five English novels by the American novelist Scott Fitzgerald: *"This Side of Paradise"* (1920), *"The Beautiful and Damned"* (1922), *"The Great Gatsby"* (1925), *"Tender Is the Night"* (1934), and *"The Last Tycoon"* (1941).

The features of selection are important to be mentioned. The five novels are selected for a variety of reasons. First, they are all in the same genre: fiction. Second, they are all written by a native novelist; therefore, they have the same personal linguistic background. This is because lexical luxury value does not adequately describe linguistic features very well if one wishes to compare groups with different linguistic proficiency with one another (Lauren, 2002). This united background poses no difficulty in measuring lexical luxury. Third, they have similar stylistic peculiarities, which reduces the impacts of sampling bias and leads to "concern with a unified authorial effort and consistent style" (Atkins et al., 1992, p. 2).

Biber's (1993) strategy is adopted in segmenting the corpus data to obtain representative and valid text samples. The text samples are approximately (roughly 5,000) tokens in size. The samples of a text should contain (5,000) tokens for each sample since grammatical terms are the most "frequently occurring" linguistic units (Meyer, 2004, p. 39). The grammatical items are considered to be the "top 50-word types in English" (Bloomer et al., 2005, p. 140). However, lexical items are "infrequently occurring" linguistic items (Meyer, 2004, p. 39). Large or long texts that exceed 1000 tokens are thus statistically and linguistically necessary to achieve a reliable analysis.

(c). Tools of Analysis

Many electronic software applications are used to perform linguistic analyses based on corpora, among these applications, (Word Counter Tools) *(see www.wordcounter.net)*, (WordSmith Tools), and (Microsoft Office Excel) are employed in the present study.

B. Corpus Analysis Procedures

Following the same methodology applied in Hussein (2014), the corpus data is analyzed in the same way but with slight differences. The differences lie in the uniqueness of the forms, numbers, and contents of the digital texts.

C. Reliability of Data Source

Data reliability is achieved through authentic investigation. The researchers began by downloading text corpora from the most authentic and trustworthy sources. After ensuring that these digital texts are precise copies of the originals, and correspond to the original paper-packed ones, they are divided into (83) samples, with each sample containing around (5,000) tokens, non-authorial materials such as titles, websites, and authorship descriptions are eliminated to minimize any statistical threat and to obtain the basic content of the text. The samples are converted from Portable Document Format (PDF) to Plain Text Format (TXT) so that the easy-to-use tools and programs like (WordCounter Tools) and (WordSmith Tools) can perform their function correctly.

V. Analysis and Discussion

The analysis procedures mentioned in Section Four are put into practice in this section. This Section, as a statistical analysis, goes through specific statistical analyses that might be employed as a type of quantitative demonstration developed for the five novels included in the corpus under consideration. Furthermore, the results of such a demonstration are explored and discussed later on.

A. Lexical Luxury Identification

Before conducting a corpus-based analysis, the number of tokens and lexical items used in the corpus must be determined using Ure's method of calculating LL as a percentage. According to Ure's method, the LL is calculated by dividing the amount or the number of lexical words by the number of whole words or tokens multiplied by (100). As an example for this process, the sentence (*This arrangement that the lexical development or regression of a writer's style may be measured in the context of these five novels*) has a total of (22) tokens but only (10) lexical items (arrangement, lexical, development, regression, writer, style, measured, context, novels). As a result, the LL of this sentence is ($\frac{10}{22}$x 100= 45%). The novels are chronologically arranged and analyzed according to their publication dates. This arrangement ensures that the lexical development and/or regression of Scott Fitzgerald's style may be measured in the context of these five works. The LL can be simply calculated by bringing out the number of tokens and lexical items using WordSmith Tools (8.0), as indicated in the subsections ahead.

(a). Lexical Luxury Identification of This Side of Paradise

When processing the samples separately using WS Tools, the wordlist is used to calculate the number of tokens. After removing all function words from each wordlist, the number of lexical items is manually calculated. The entire textual samples were converted to (plain text type) format (txt), then analyzed using WordSmith Tools to determine the frequency of both lexical and functional items. Following the manual calculation of the number of lexical items from

the Wordlist, the numbers of tokens as well as lexical items are determined using Ure's formula to provide LL percentages. Throughout the five novels, the same techniques are employed. The example in Table 2 comes from sample number (1) from "*This Side of Paradise*".

TABLE 2
TOKENS, LEXICAL ITEMS, AND LL PERCENTAGES OF (16) TEXTUAL SAMPLES OF *"THIS SIDE OF PARADISE"*

Sample No.	Tokens	Lexical Items	LL.
1	5.006	1.729	34.54 %
2	5.006	1.648	32.92 %
3	5.008	1.779	35.52 %
4	5.002	1.562	31.23 %
5	5.006	1.627	32.50 %
6	5.000	1.519	30.38 %
7	5.004	1.608	32.13 %
8	5.003	1.514	30.26 %
9	5.002	1.509	30.17 %
10	5.005	1.499	29.95 %
11	5.007	1.266	25.28 %
12	5.001	1.666	33.31 %
13	5.001	1.583	31.65 %
14	5.000	1.533	30.66 %
15	5.006	1.762	35.20 %
16	4.699	1.466	31.20 %
Average	**79.756**	**25.270**	**31.68 %**

In the statistical description in Table 2, the whole novel (*This Side of Paradise*) is segmented sequentially into (16) samples. Approximately, each sample of text holds approximately (5000) tokens. The LL scale for these samples ranges from (25.28%) the low percentage to (35.52 %), the highest percentage. The weight of content or lexical items varies across samples, despite the fact that their frequencies are generally convergent across all samples. Samples with a high frequency of lexical items score high LL percentages, as with the third sample. The third sample scores (35.52 %) LL percentage with (5.008) tokens and (1.779) lexical words. The samples of lower-frequency lexical words score low LL percentages. The eleventh sample has the lowest LL rate (25.28 %) with (5.007) tokens and (1.266) lexical words. Figure 1 shows the percentages of this LL scale.

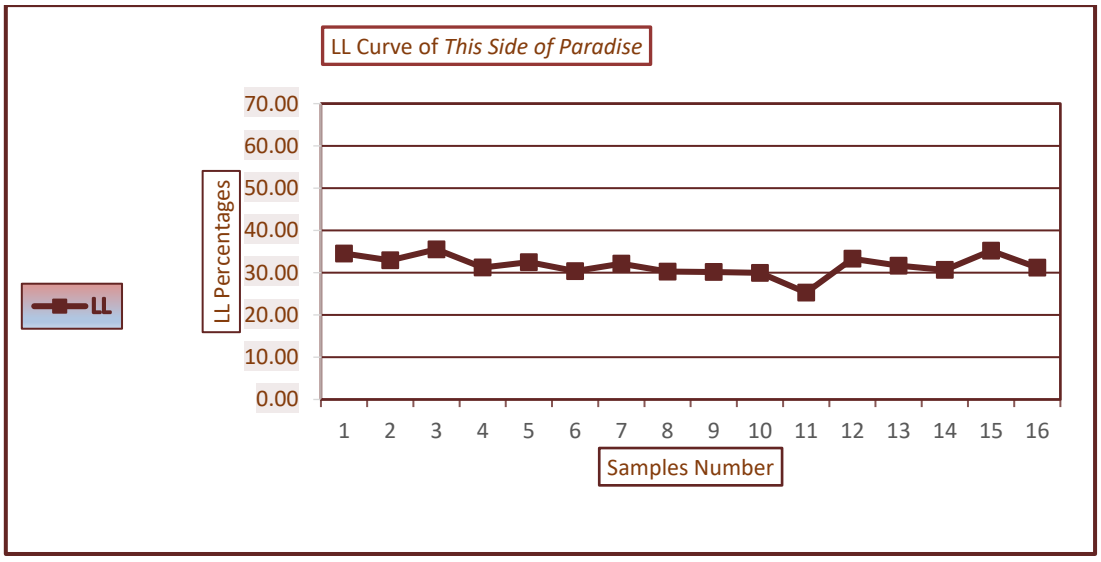

Figure 1. The Order of LL Percentages of (16) Textual Samples of "*This Side of Paradise*"

The LL curve in Figure 1 visualizes convergent LL percentages. These differences among LL ratios push this curve stretch forward in a rather instable progression. It is noticeable on this curve that its peak occurs at the third sample (35.52 %) with (5.008) tokens and (1.779) lexical items.

(b). Lexical Luxury Identification of "The Beautiful and Damned"

The next table introduces the statistical description of lexical items, tokens, and their LL percentages of (25) samples visualizing the whole textual material of *"The Beautiful and Damned"*:

TABLE 3
TOKENS, LEXICAL ITEMS, AND LL PERCENTAGES OF (25) TEXTUAL SAMPLES OF "*THE BEAUTIFUL AND DAMNED*"

Sample No.	Tokens	Lexical Items	LL.
1	5.003	1.732	34.62 %
2	5.002	1.625	32.49 %
3	5.005	1.635	32.67 %
4	5.000	1.634	32.68 %
5	5.002	1.563	31.25 %
6	5.007	1.652	32.99 %
7	5.001	1.500	29.99 %
8	5.009	1.631	32.56 %
9	5.006	1.598	31.92 %
10	5.003	1.582	31.62 %
11	5.008	1.668	33.31 %
12	5.008	1.575	31.45 %
13	5.005	1.628	32.53 %
14	5.002	1.535	30.69 %
15	5.003	1.633	32.64 %
16	5.002	1.608	32.15 %
17	5.007	1.678	33.51 %
18	5.003	1.715	34.28 %
19	5.008	1.535	30.65 %
20	5.003	1.653	33.04 %
21	5.004	1.513	30.24 %
22	5.002	1.641	32.81 %
23	5.004	1.555	31.08 %
24	5.002	1.433	28.65 %
25	3.380	1.137	33.64 %
Average	**123.479**	**39.659**	**32.14 %**

As presented in Table 3, the (25) samples of the second novel, "*The Beautiful and Damned*", present distinct LL percentages. The LL scale starts consisting of (28.65%) as the lowest percentage and ends up with (34.62%) as the highest one. The first sample, on the one hand, scores (34.62%) with (5,003) tokens and (1.732) lexical items. The last sample, on the other hand, scores (33.64%) with (3.380) tokens and (1.137) lexical items. Samples (2, 5, 14, 16, 22, 24) and Samples (1, 10, 15, 18, 20) consist of a similar number of tokens, nonetheless they score different LL percentages. The numbers of the various lexical items make the LL percentages for the samples marked up successively with (32.49%) having (1.625) lexical items, (31.25%) with (1.563) lexical items, (30.69%) with (1.535) and (32.15%) with (1.608) lexical items, (32.81%) with (1.641), (28.65%) with (1.433) lexical items. The curve of this scale is plotted visually in Figure 2:

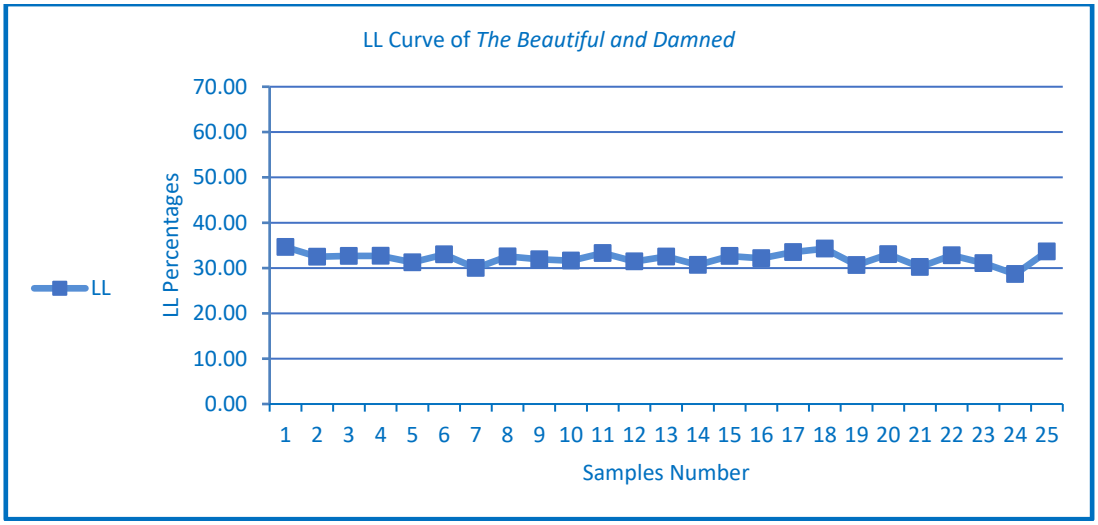

Figure 2. The Order of LL Percentages of (25) Textual Samples of "*The Beautiful and Damned*"

The curve of the LL scale of *"The Beautiful and Damned*" starts with (34.58%) LL ratio, and ends with roughly similar LL ratio (33.64%). Furthermore, the lowest point of this curve is found at the twenty-eight-sample scoring (28.41%) LL percentage. This LL percentage is the lowest one among the overall percentages of this novel.

(c). Lexical Luxury Identification of "The Great Gatsby"

After the sequential sampling method being applied, the whole textual material of *"The Great Gatsby"* is divided into (10) samples. The samples hold approximately (5000) tokens for each sample. Then the samples are processed

separately with (WS) Tools. The same procedures and processes followed in processing the samples of the last novels are applied to the (10) samples of *"The Great Gatsby*".

The table demonstrates the number of tokens with lexical items and LL percentages of the (10) samples of *"The Great Gatsby*":

TABLE 4
TOKENS, LEXICAL ITEMS, AND LL OF (10) TEXTUAL SAMPLES OF *"THE GREAT GATSBY"*

Sample No.	Tokens	Lexical Items	LL.
1	5.026	1.709	34.00 %
2	5.020	1.876	37.37 %
3	5.002	1.804	36.06 %
4	5.006	1.856	37.07 %
5	5.008	1.598	37.06 %
6	5.005	1.798	35.92 %
7	5.000	1.800	36.00 %
8	5.011	1.776	35.44 %
9	5.001	1.691	33.81 %
10	2.803	1.009	35.99 %
Average	**47.882**	**16.917**	**35.87 %**

As Table 4 presents, the LL scale of the (10) samples of *"The Great Gatsby"* reveals various percentages. Sample (2) holds the highest level of percentage on this scale: (37.37%) with (1.876) lexical items, and (5.020) tokens. However, the sample (8) holds the lowest level of percentage (33.81%) with (1.691) lexical items and (5,001) tokens. The rest of the samples had nearly identical percentages, ranging from (34.00%) in the 1st sample to (37.07%) in the 4th sample. It suggests that lexical items play a significant role in eliciting varying LL percentages. Consequently, the LL percentage is calculated using the amount or the number of lexical items. Thus, a large number of lexical Thus, a large number of lexical items results in a high LL percentage, whereas a small number of lexical items results in a low LL percentage. The next graph depicts the LL curve of the (10) samples of *"The Great Gatsby*".

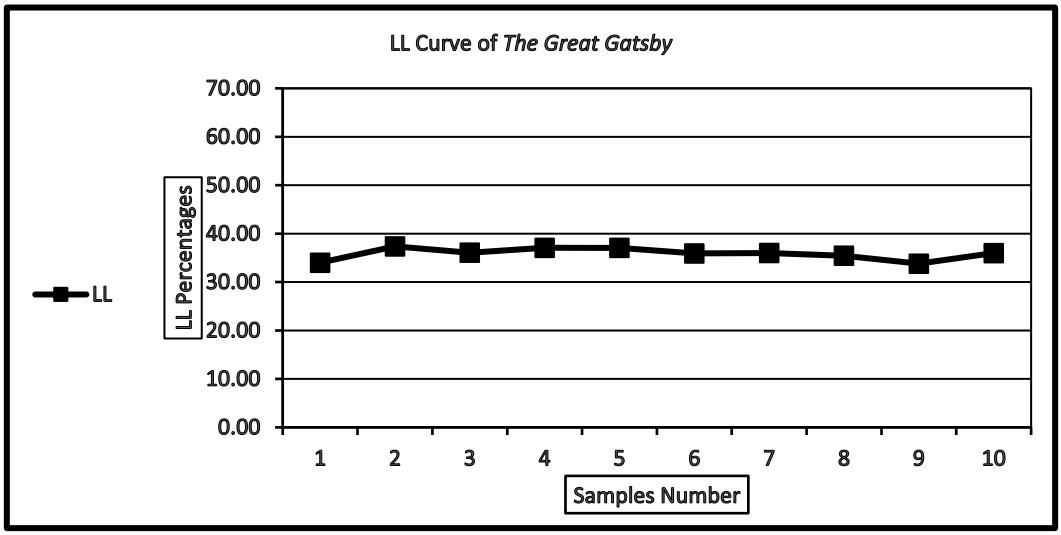

Figure 3. The Order of LL Percentages of (10) Textual Samples of *"The Great Gatsby*"

As clearly shown, the curve in the figure above presents the statistical behavior of LL percentages. The percentage of the highest LL (37.37%) of the second sample marks the peak of this curve, while the ninth sample's lowest LL percentage (33.81%) indicates the low point.

(d). Lexical Luxury Identification of "Tender Is the Night"

The *wordlist* is used to calculate the number of tokens while processing the samples using *WS Tools* for *"Tender is The Night"*. The number of lexical items when calculated manually after deleting all function words from each wordlist is (38,995) as the total of (22) samples of *Tender is The Night*, whereas the total number of tokens is (108.373). After manually calculating the lexical words and tokens from the *Wordlist*, the tokens, and lexical item numbers are calculated using Ure's method to provide LL percentages. Similar strategies are used throughout the fourth novel.

Table 5 displays the statistical overview of tokens, lexical items and LL percentages of (22) samples of *"Tender is the Night":*

TABLE 5
TOKENS, LEXICAL ITEMS AND LL PERCENTAGES OF (22) TEXTUAL SAMPLES OF *"TENDER IS THE NIGHT"*

Sample No.	Tokens	Lexical Items	LL.
1	5.003	1,786	35.72 %
2	5.001	1,791	35.97 %
3	5.002	1,798	35.93 %
4	5.006	1,801	35.99 %
5	5.002	1,837	36.73 %
6	5.018	1,869	37.35 %
7	5.017	1,852	37.02 %
8	5.028	1,821	36.39 %
9	5.020	1,803	36.03 %
10	5.006	1,789	35.71 %
11	5.047	1,785	35.67 %
12	5.000	1,776	35.51 %
13	5.012	1,759	35.14 %
14	5.029	1,812	36.18 %
15	5.009	1,807	36.12 %
16	5.005	1,731	36.51 %
17	5.007	1,800	33.95 %
18	5.003	1,873	37.42 %
19	5.008	1,798	35.95 %
20	5.032	1,826	36.43 %
21	5.011	1,783	35.60 %
22	3.107	1,098	35.76 %
Average	**108.373**	**38,995**	**36.05 %**

As shown in Table 5, the (22) samples of the fourth novel, *"Tender is The Night"*, depict different LL percentages. The LL value starts with (33.95 %) as the lowest ratio, ending up with (37.42 %) as the highest ratio. Meanwhile, the 1st sample gets (35.72 %) with (5,003) tokens and (1.786) lexical items. However, the last of these sample scores (35.76 %) with (3.107) tokens and (1.098) lexical items. What appears to be fascinating is that the number of lexical items plays a vital role in triggering varied LL percentages. As a result, the LL percentage is heavily influenced by the quantity of lexical items. A high number of lexical items inevitably results in a high LL percentage, whereas a low number of lexical items results in a low LL percentage. The graph below is a graphic depiction of the LL curve of the (22) samples of *"Tender is the Night"*:

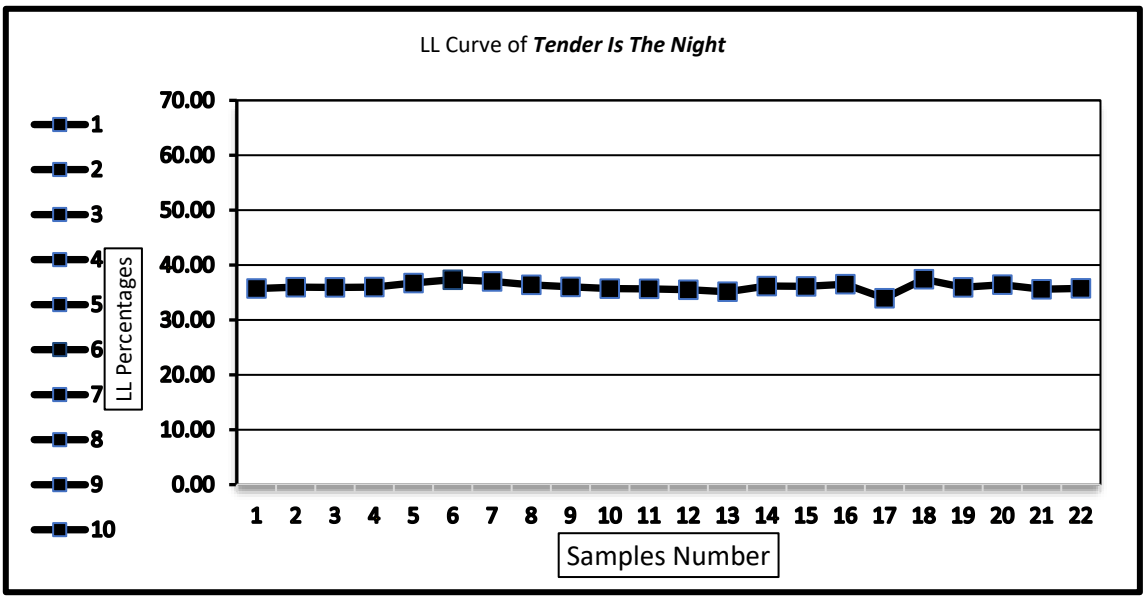

Figure 4. The Order of LL Percentages of (22) Textual Samples of *"Tender Is the Night"*

The curve depicted in the figure above, as clearly shown, depicts the statistical behavior of LL percentages. The peak of this curve is represented by the highest percentage of LL (37.42 %) of the eighteenth sample, while the falling point is represented by the lowest percentage of LL (33.95 %) of the seventeenth sample.

(e). Lexical Luxury Identification of "The Last Tycoon"

Table 6 visualizes the tokens, lexical items, and LL percentages of the (10) samples of *"The Last Tycoon"*:

TABLE 6
TOKENS, LEXICAL ITEMS, AND LL OF (10) TEXTUAL SAMPLES OF *"THE LAST TYCOON"*

Sample No.	Tokens	Lexical Items	LL.
1	5.001	1,801	36.01%
2	5.006	1,798	35.91 %
3	5.000	1,819	36.38 %
4	5.002	1,789	35.76 %
5	5.005	1,793	36.18 %
6	5.008	1,790	35.82 %
7	5.004	1,806	35.09 %
8	5.007	1,822	35.86 %
9	5.003	1,801	36.41 %
10	1.966	786	39.97 %
Average	**47.002**	**17.005**	**36.34 %**

As seen in Table 6, the LL scale of the (10) samples displays surprisingly varied percentages. The tenth sample has the highest ratio (39.97 %) on this scale, with (786) lexical items and (1.966) tokens. The sample (7) has the lowest percentage (35.09 %) with (1,806) lexical items and (5.004) tokens. The rest of the samples had nearly identical percentages, ranging from (35.76 %) in the fourth sample to (36.41 %) in the ninth sample. What appears to be fascinating is that the number of lexical items plays a big role in triggering fluctuated LL percentages. As a result, the LL percentage is heavily influenced by the amount of lexical items.

A high number of lexical words inevitably results in a high LL percentage, whereas a low lexical item number results in a low LL percentage. The LL curve of the (10) samples of *"The Last Tycoon"* is represented visually in the graph below.

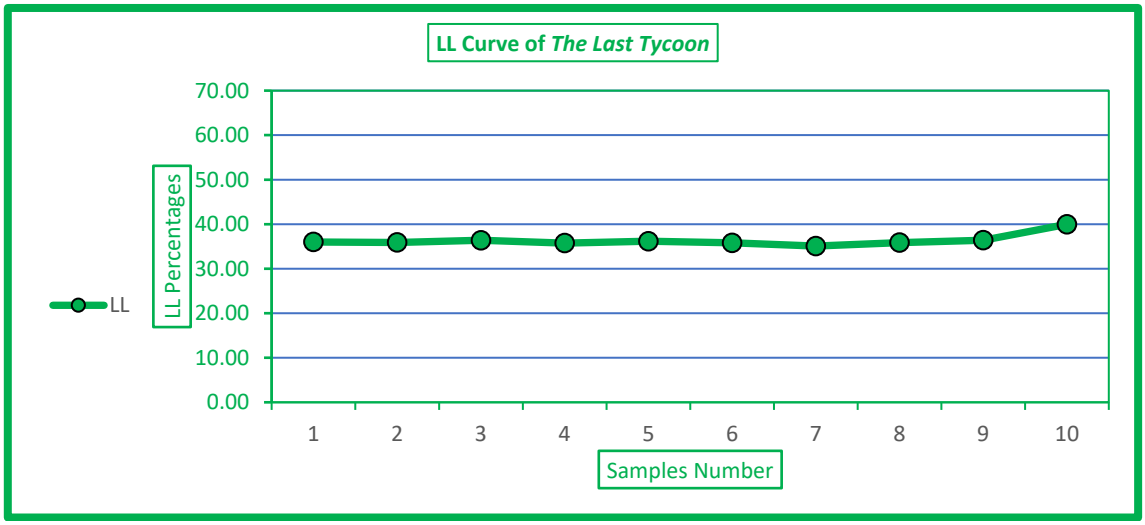

Figure 5. The Order of LL Percentages of (10) Textual Samples of *"The Last Tycoon"*

"The Last Tycoon's" LL scale curve, as depicted in the graph above, begins with (36.01%) LL percentage and ends with different LL percentages (39.97 %). Furthermore, the curve's lowest point occurs at the seventh sample scoring (35.09 %) LL percentage, this LL percentage is the lowest of the novel's overall percentages.

B. Discussion

The next step is to obtain the results of these statistical outputs after constructing the quantitative description of LL identification of the corpus. The preceding LL identification is the mathematical process of establishing and identifying LL percentages using a statistical formula. This is the first step in creating a lexical luxury analysis. Lexical luxury is a graphical representation that shows how much each novel displays lexical characteristics or features as defined by the LL percentages.

Then, it may be rather useful to compare the LL curves of the five novels in order to make a direct and observable comparison. The LL is reflected in the table below by the LL percentages of the five novels.

TABLE 7
LL OF *"THIS SIDE OF PARADISE"* (1920), *"THE BEAUTIFUL AND DAMNED"* (1922), *"THE GREAT GATSBY"* (1925), *"TENDER IS THE NIGHT"* (1934), AND *"THE LAST TYCOON"* (1941)

Samples No.	This Side of Paradise	The Beautiful and Damned	The Great Gatsby	Tender Is The Night	The Last Tycoon
1	34.54 %	34.62 %	34.00 %	35.72 %	36.01%
2	32.92 %	32.49 %	37.37 %	35.97 %	35.91 %
3	35.52 %	32.67 %	36.06 %	35.93 %	36.38 %
4	31.23 %	32.68 %	37.07 %	35.99 %	35.76 %
5	32.50 %	31.25 %	37.06 %	36.73 %	36.18 %
6	30.38 %	32.99 %	35.92 %	37.35 %	35.82 %
7	32.13 %	29.99 %	36.00 %	37.02 %	35.09 %
8	30.26 %	32.56 %	35.44 %	36.39 %	35.86 %
9	30.17 %	31.92 %	33.81 %	36.03 %	36.41 %
10	29.95 %	31.62 %	35.99 %	35.71 %	39.97 %
11	25.28 %	33.31 %		35.67 %	
12	33.31 %	31.45 %		35.51 %	
13	31.65 %	32.53 %		35.14 %	
14	30.66 %	30.69 %		36.18 %	
15	35.20 %	32.64 %		36.12 %	
16	31.20 %	32.15 %		36.51 %	
17		33.51 %		33.95 %	
18		34.28 %		37.42 %	
19		30.65 %		35.95 %	
20		33.04 %		36.43 %	
21		30.24 %		35.60 %	
22		32.81 %		35.76 %	
23		31.08 %			
24		28.65 %			
25		33.64 %			
Average	**31.68 %**	**32.14 %**	**35.87 %**	**36.05 %**	**36.34 %**

The LL percentages are quantitatively different, with each profile displaying diverse ratios on its scale. Aside from the differences in LL ratios or percentages, these profiles or columns differ in size, for example, in terms of the number of samples. *"The Last Tycoon*" has the highest LL average (36.34 %). While the other four novels have profiles that are quite close to the LL average, (31.68%) for *"This Side of Paradise"*, (32.14%) for *"The Beautiful and Damned*", (35.87) for *"The Great Gatsby*", and (36.05%) for *"Tender Is The Night*". The five profiles must be graphically plotted together in one graph for a more accurate and significant visual representation. As a result of this visual representation, the LL of each novel would be clearly noticed and compared.

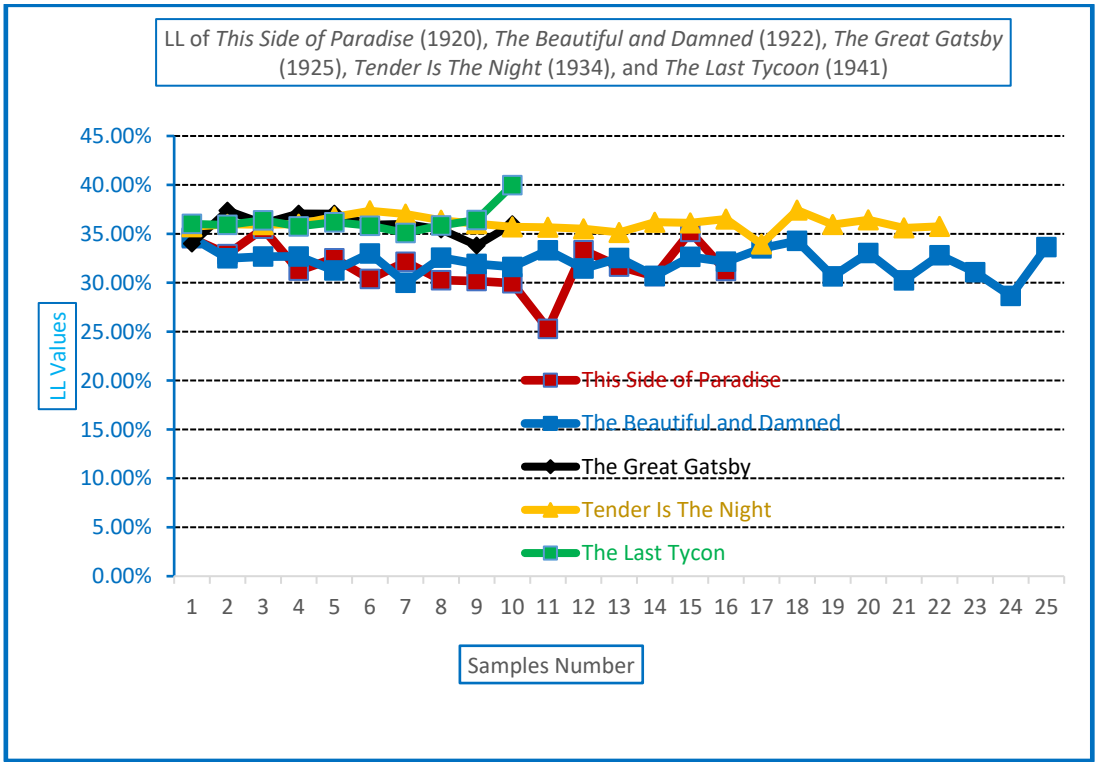

Figure 6. The Statistical Behavior of LL curves of: *"This Side of Paradise" (1920), "The Beautiful and Damned" (1922), "The Great Gatsby" (1925), "Tender Is the Night" (1934), and "The Last Tycoon" (1941).*

Following the curves shown in Figure 4, 5, a notable difference in the LL percentages that comprise the LL of the five novels can be noted. Neumann (2014) states that;

> A high lexical density can be interpreted as an indirect indicator of frequent lexical reference since a high proportion of lexical words at the expense of function words, pronouns in particular, reflects the use of explicitly verbal rather than other resources to refer to objects or concepts. (p. 159)

According to the quantitative method in her study, Neumann (2014) claims that lexical luxury is a more appropriate measure for conducting an "investigation of orientation towards content" (p. 159) because it focuses on content words. As a result, the differences between the LL curves in the graphic above are due to vocabulary loads. In comparison to the other four novels, the percentages of LL in "*The Last Tycoon*" are the highest. It continues to rise from (35.09%) to (39.97 %). With (47.002) tokens and (17,005) lexical items, this novel has the highest LL ratio (36.34 %).

The curves of *"This Side of Paradise", "The Beautiful and Damned",* and *"The Great Gatsby*" begin with almost the same values in the first samples (34.54%), (34.62 %), and (34.00 %). They, on one hand, diverge directly in the second sample where "*This Side of Paradise*" achieves a score of (32.92 %), "*The Beautiful and Damned*" scores (32.49 %), "*The Great Gatsby*" gets a score of (37.37 %), and "*Tender Is The Night*" has a score of (35.97 %).

On the other hand, they end up with different LL values where "*This Side of Paradise*" scores the lowest LL percentage. It ends with (31.20 %) and (31.86%) as a total LL ratio with (79.756) tokens and (25.270) lexical items. "*The Beautiful and Damned*" scores a higher LL value than "*This Side of Paradise*" at its total ratio (32.14 %) with (123.479) tokens and (39.659) lexical items.

It is worth noting that although "*The Great Gatsby*" is similar to "*The Last Tycoon*" in terms of the number of samples (10 for both), they are not similar in terms of percentages or number of lexical words. The main reason seems to be the long period between the two novels, which in turn leads to noticeable differences in the writer's linguistic ability.

Despite the high LL percentages of each sample, the "*Tender Is The Night*" curve appears to be more lexically denser or more luxurious than "*The Great Gatsby*", "*The Beautiful and Damned*", and "*This Side of Paradise*" because the overall average of its lexical items is more than the total average of the other three novels. The quantitative analysis of the five novels is based on lexical luxury averages and percentages. "The Last Tycoon's" lexical luxury averages are quantitatively higher than those of "*This Side of Paradise*", "*The Beautiful and Damned*", "*The Great Gatsby*", and "*Tender Is the Night*". As a result, "*The Last Tycoon*" takes the first position with the highest LL average. "*Tender Is The Night*" comes at second place as being lexically more luxurious than "*The Great Gatsby*", "*The Beautiful and Damned*", and "*This Side of Paradise*" which comes at fifth place.

VI. Conclusions

To round off this study, the rapid growth or development of corpus linguistics in general, and corpus stylistics in particular, makes it possible to collect, classify, segment, analyze, and monitor large amounts of linguistic data. Corpus-based stylistic studies have proven to be effective in exploring specific aspects of style. However, the combination of corpus linguistics and stylistics (i.e., corpus stylistics) makes significant use of statistical descriptions offered by trustworthy and authentic measures such as Type/Token Ratio. In other words, corpus stylistics has proven to be extremely fruitful in investigating the statistical behavior presented by TTR gauges of lexical resources within texts involved in the analysis process.

Corpus linguistics, which is based on machine-readable databases, provides some basic feasible techniques for obtaining and analyzing data, building corpora, developing new models and tools of analysis, as well as answering research questions. In addition, it shows a collection of statistical procedures which can be applied to text analysis. In addition, linguistic characteristics such as lexical luxury can be measured using statistical approaches. The statistical and stylistic analyses are carried out under the umbrella of corpus stylistics, where there is genuine collaboration and real cooperation between corpus linguistics and stylistics. This is consistent and goes with the first hypothesis, which states that corpus-based techniques can be highly useful and helpful in producing empirical and quantitative stylistic descriptions of literary works in terms of lexical luxury.

The prior statistical step for developing lexical luxury profiles is lexical luxury identification using corpus linguistic methods. The lexical luxury profile is an adequate graphical representation for identifying lexicality in literary works, which verifies the second hypothesis as an efficient measure for lexicality and lexical development of the writer's style. This study's corpus consists of five novels published by Scott Fitzgerald, the lengths of these novels vary. The lexical luxury curves of these novels in the corpus demonstrated varying percentages of lexical luxury with (5000) tokens and (1000-2000) lexical items in each sample.

As the third hypothesis previously assumed, the high percentages of lexical luxury are due to high concentrations of lexical items resulting in luxurious lexical profiles. As a result, the values of the lexical luxury profile are sensitive to the number of lexical items. The analysis of the five novels using WordSmith Tools reveals various lexical luxury averages and percentages. In its lexical luxury profile, The Last Tycoon has higher lexical luxury percentages as well as a distinct lexical luxury average. *Tender is the Night* ranks second in terms of lexical luxury, then respectively: *The Great Gatsby*, *The Beautiful and Damned*, and *This Side of Paradise* which is in fifth place.

Furthermore, these five novels are studied and analyzed chronologically based on their original publication dates. This chronological arrangement of the corpus texts demonstrates that Scott Fitzgerald's linguistic choices get increasingly sophisticated over the context of these five novels. The high LL averages of each novel score, (1) *The Last Tycoon* (36.34%), (2) *Tender Is The Night* (36.05%), (3) *The Great Gatsby* (35.87), *The Beautiful and Damned* (32.14%), and *This Side of Paradise* (31.68%), they constitute quantitative evidence that confirms interesting progress in Fitzgerald's lexical performance. Additionally, because they have high LL percentages, such percentages indicate model averages for written fictional genres. These percentages exceeded the written text lexical luxury ratio of (30%).

After all the high frequency of lexical words in Fitzgerald's writing style is the linguistic trait that distinguishes his style as dense style. Thus, lexical luxury is the salient feature that has been quantitatively measured. Then it is obvious that lexical luxury could be regarded as a 'stylistic marker' in literary texts. It can reveal much about the style and how ideas and information are delivered in the text where the lexical choices can be clearly identified.

These choices are represented in higher lexical luxury percentages and denser lexical choices in which factual information is used to convey events in words. In terms of Fitzgerald's style, LL quantitatively demonstrates that its distinct solidity is found in precise sentences. This precise writing style contains a variety of stylistically diverse terms as well as obvious references to nouns and concepts. As a result of Fitzgerald's five novels, luxurious linguistic styles emerged. Fitzgerald's early journalist training and constant travel may have given his lexicon its depth and uniqueness, which are practically represented in his writing style.

REFERENCES

[1] Al-Wahy, A. (2016). " Towards a Methodology for Measuring Lexical Density in Arabic". *The International Journal of Arabic Linguistics (IJAL), Vol.3*, Issue 1, pp. 1-33.

[2] Anthony, L. (2009). "Issues in the Design and the Development of Software Tools. 2009. for Corpus Studies: The Case for Collaboration". In Baker, P. 2012. (ed.) *Contemporary Corpus Linguistics*. London: Continuum. International Publishing Group.

[3] Atkins, S., Clear, J. and Ostler, N. (1992). "Corpus Design Criteria". *Literary and Linguistic Computing*, *7*(1), 1-16.

[4] Biber, D. (1993). "Representativeness in Corpus Design". *Literary and Linguistic Computing*, *Vol. 8*, No. 4, 1993. Oxford University Press.

[5] Biber, D., & Reppen, R. (2015). "Introduction". In Biber, D. & Reppen, R. (eds.), 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

[6] Bloomer, A., Griffths, P. and Merrison, A. (2005). *Introducing Language in Use: A Coursebook*. London & New York: Routledge.

[7] Camiciottoli, B. (2007). *"The Language of Business Studies Lectures: A Corpus-assisted Analysis*. Amsterdam & Philadelphia: John Benjamins Publishing Company".

[8] Eder, M. (2011). "*Style-Markers in Authorship Attribution: A Cross-Language Study of the Authorial Fingerprint*". Pedagogical University of Kraków. Polish Academy of Sciences, Institute of Polish Language.

[9] Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics*. (2nd ed.). New York & London: Continuum.

[10] Enkvist, N. (1973). *Linguistic Stylistics*. Paris: Mouton.

[11] Fitzgerald, F. Scott. (1925). *The Great Gatsby*. New York: C. Scribner's E-Book, Gutenberg Australia.

[12] Fitzgerald, F. Scott. (1945, 2009). Wilson, Edmund (ed.). *The Crack-Up*, New Directions, New York, USA.

[13] Fitzgerald, Scott. (1920, 2018). *This Side of Paradise*. Gutenberg E-Book, Gutenberg Australia.

[14] Fitzgerald, Scott. (1922, 2020). *The Beautiful and Damned*. Gutenberg E-Book, Gutenberg Australia.

[15] Fitzgerald, Scott. (1934, 2021). *Tender is The Night*. Gutenberg E-Book, Gutenberg Australia.

[16] Fitzgerald, Scott. (1941, 2003). *The Last Tycoon*. Gutenberg E-Book, Gutenberg Australia.

[17] Francis, N. (2017). *Bilingual and Multicultural Perspectives on Poetry, Music, and Narrative: The Science of Art*. USA: Lexington Books.

[18] Giannossa, L. (2012). "*A Corpus-Based Investigation of Lexical Cohesion in EN and IT Non-Translated Texts*". Unpublished Ph.D. Dissertation. Kent State University.

[19] Halliday, M. (1989). *Spoken and Written Language*. Oxford: Oxford University Press.

[20] Hickey, L. (1993). "Stylistics, Pragmatics and Pragmastylistics". In *Revue belge de philologie et d'histoire*, 573-86.

[21] Holmes, D. (1994). "Authorship Attribution". *Computers and the Humanities*, *28*, 87-106, 1994. Netherlands: Springer.

[22] Hunston, S. (2022). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

[23] Hussein, K. (2014). *A Computational Linguistics Perspective of Authorship and Plagiarism*. Berlin: Lambert Academic Publishing.

[24] Johansson, V. (2008). "*Lexical Diversity and Lexical density in Speech and Writing: A developmental Perspective*" Working Papers. Lund University, Dep. of Linguistics & Phonetics, *53*(2008), 61-76.

[25] Keune, K. (2012). "*Explaining Register and Sociolinguistic Variation in the Lexicon: Corpus Studies on Dutch*". Ph.D. Dissertation Published by LOT. Netherlands: Trans, 10, 3512 JK Utrecht.

[26] Lauren, U. (2002). "*Some lexical features of immersion Pupils' Oral and Written Narration*". Lund University, Dep. of Linguistics, Working Papers, *50*(2002), 63-78.

[27] McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: An Introduction* (2nd ed.). Edinburgh: Edinburgh University Press.

[28] McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

[29] McIntyre, D. (2012). "Corpus Stylistics in the Classroom". In Burke, M., Csabi, S., Week, L. & Zerkowitz, J. (eds.) 2012. *Pedagogical Stylistics: Current Trends in Language, Literature and ELT*. London: Bloomsbury Publishing.

[30] McIntyre, D. (2015). "Towards an Integrated Corpus Stylistics". *Topics in Linguistics*. Issue 16. University of Huddersfield, United Kingdom.
[31] McMenamin, G. (2002). "Stylistics". In McMenamin, G., Choi, D., Mistry, P., Morton, S. & Yasuda, W. (eds.). *Forensic Linguistics: Advances in Forensic Stylistics*. USA: CRC Press.
[32] Meyer, C. (2004). *English Corpus Linguistic: An Introduction*. Cambridge: Cambridge University Press.
[33] Neumann, S. (2014). *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. Germany: DE GRUYTER Mouton.
[34] Norgaard, N., Montero, R., & Busse, B. (2010). *Key Terms in Stylistics*. London: Continuum International Group.
[35] Perfetti, C. (2000). "Lexical Density and Phrase Structure Depth as Variables in Sentence Retention". *Journal of Verbal Learning and Verbal Behavior*, *8*, 719-724. (1969). Bell & Howell Information & Learning Company.
[36] Shepherd, T. & Sardinda, T. (2013). *"A Rough Guide to Doing Corpus Stylistics"*. Matraga, *Rio De Janeiro*, *20*, (32).
[37] Simpson, P. (2004). *Stylistics: A Resource Book for Students*. London: Routledge.
[38] Singla, A. (2012). *"The Relationship between Lexicon and Syntax in Texts Written in Catalan by School Children and Adolescents*". Ph.D. Dissertation Published by (www.TDX.com.). University of De Barcelona.
[39] Smith, J. & Kelly, Colleen. (2002). Stylistic Constancy and Change Across Literary Corpora: Using Measures of Lexical Richness to Date Works. *Computers and the Humanities*, *36*, 411-430. 10.1023/A:1020201615753.
[40] Ure, J. (1971). Lexical Density and Register Differentiation. *Contemporary Educational Psychology*, *5*, 96-104.
[41] Vera, G., Sotomayor, C., Bedwell, P., Dominguez, A. & Jelderez, E. (2016)." *Analysis of Lexical Quality and Its Relation to Writing Quality For 4th Grade, Primary School Students in Chile*". Centre of Advanced Research in Education University of Chile. Santiago: Springer.
[42] Verdonk, P. (2002). *Stylistics*. Oxford: Oxford University Press.
[43] Yeibo, E. (2011). "Group Types as Style Markers in J.P. Clark-Bekeremo's Poetry". *Theory and Practice in Language Studies*, *Vol. 8*, pp. 92-934, August 2011. Finland: Academy Publisher.
[44] Yumin, C. (1986). "An Attempt at Analyzing Linguistic Style". In Aarts, J. & Meijs, W. (eds.). (1986). *Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora*. Netherlands.

Hasan Ali Husein was born in Iraq in 1991. He is a PhD. Candidate in English and Linguistic studies in the Department of English – College of Education for Human Sciences- University of Babylon and a Lecturer at Nasirryah Technical Institute in Southern Technical University. He received his master's degree in English from Thi-Qar University, Iraq in 2021. He has presented numerous research papers at international Journals. His main research and academic interests include English for Specific Purposes, Language Skills Development, Contrastive Linguistics, and Stylistics.

Nesaem M. Al-Aadili was born in Babylon/ Iraq on the fourteenth of July 1983. B.A in the English language in 2005, MA in English language and linguistics in 2011, and PhD in English language and linguistics in 2018; the three degrees are from the University of Babylon/ Babylon/ Iraq. She works as a Professor at the University of Babylon/ College of Education for Human Sciences/ Department of English. She is interested in writing articles in the fields of pragmatics and critical discourse analysis.