International Journal of Psychosocial Rehabilitation, Vol.24, Issue 01, 2020
ISSN: 1475-7192

A Corpus-Based Stylistic Identification of Lexical Density Profile of Three Novels by Virginia Woolf: The Waves, Mrs. Dalloway and To the Lighthouse

[1] Khalid Shakir Hussein, [2] Rawa Abdul-Shareef Kadhim

ABSTRACT

This paper is designed to present a short panorama of one of the stylistic markers which is that of lexical density. This marker is to be investigated within three literary texts of Virginia Woolf using the lexical density ratio with the aims of: (i) discovering how useful is the lexical density ratio in measuring the lexical resources of Virginia Woolf's authorial style as it proceeds in her three novels: The Waves (1931), Mrs.Dalloway (1925),To the Lighthouse (1927) and (ii) comparing the lexical density ratios of these three novels so as to determine how, where and when the author's lexical repertoire has prospered and/or declined. Downloading the digital texts, they are distributed into (43) samples: Mrs.Dalloway (13) samples, To the Lighthouse (14) samples and The Waves (16) samples. Each sample is of an approximate number of (5000) tokens. These samples are processed via WordSmith Tools (4.0) so as to obtain the number of tokens and lexical words. The obtained numerical data are then transformed into visual graphs via Microsoft Excel. It is found that Woolf's The Waves seizes the majority of the higher lexical density ratios, whereas To the Lighthouse comes to score the lowest ratios. Mrs.Dalloway occupies an inbetween position where its curve stretches between that of The Waves and To the Lighthouse. Accordingly, The Waves has been proved to be the denser amongst the other two novels. A quality which gives it the shadow of being a difficult text.

Key Words: *corpus stylistics, lexical density, functional words, content words.*

INTRODUCTION

Corpus stylistics can be described as a hybrid field for it represents that meeting area between two disciplines: stylistics and corpus linguistics. Corpus stylistics makes use and employs the techniques of corpus linguistics in analyzing texts, whether literary or non-literary. Thus, it combines the linguistic descriptions with the literary appreciations. When it comes to be compared with other fields of study, corpus stylistics is relatively a new one. Consequently, only a limited number of investigations are carried out within and subsumed under the heading of corpus stylistics. Nevertheless, corpus stylistics comes to prove itself as an efficient and objective way in dealing with certain textual issues which would be otherwise subjective, critical and controversial. As one of the controversial issues is that of measuring and characterizing the writing style of a particular author. Studies of this type have been criticized of being subjective since they can be affected by the subjective judgments. Lexical density is one these stylistic issues that represents some controversy because there is no precise consensus on the fixed limits that distinguish between function and content words. Thus, to overcome bias and subjectivity and to achieve a high degree of objectivity, corpus stylistics comes to adopt corpus linguistics tools in counting and obtaining the number of such words and items putting them in the form of numerical data. These data are to be processed via some mathematical formula and so the lexical density of a particular author's style can be identified accordingly. This paper puts forward the following hypothesis: (i) the lexical density ratio can be effectively used in describing and characterizing the

Khalidshakir74@gmail.com[1] *English Department, College of Arts, Thi-Qar University, Iraq. E-mail:*
[2] *English Department, College of Education for Humanities, Thi-Qar University, Iraq.* almwswya322@gmai.com

authorial style of a given writer in a rather objective and non-biased way, (ii) time can have an effect upon the writers' authorial style, that is to say, a given author's lexical repertoire is subject to change throughout the course of time. Consequently, it is open to development or decline and (iii) the denser a given author's writing style is, the more complicated and richer his/her texts tend to be.

Lexical Density and Text Complexity

Generally, two main types of forms can be subsumed under the heading of lexical items. The first type is represented by forms like nouns, adjectives, adverbs and verbs, which is an open class, and the second is represented by function forms such as pronouns, articles, prepositions, etc, referred to as a closed system. More specifically, lexical density is more associated with the former type of words than the latter. In this sense, lexical density can work as a measure of lexical development (Halliday,1990: 61, & Kondal,2015: 25). Halliday (ibid: 62) points out that the written form of language makes use of more lexical items than the spoken form and thus, it is considered as denser. In his exact words, "written language is dense, spoken language is sparse".

Since lexical density is mainly associated with the content forms rather than the functional ones, it is useful to draw a precise distinction between the two types. Content forms are those words that carry the gist of the message, that is to say, their informational load is high. They are called content words. They include nouns, adjectives, verbs and adverbs (Yule,2010: 68, & Baker et al.,2006: 47). Functional forms, on the other hand, are those words that have some grammatical functions in a given structure, that is why they are also called grammatical words. They include prepositions, articles, pronouns, conjunctions, etc. When measuring the lexical density of a given text, functional forms are usually excluded (Yule: ibid, & Baker et al.,2006: 76).

Ure (1971: 445) presents a distinction between the previously mentioned types in terms of their lexical properties. Thus, forms with no lexical properties can be characterized as pure grammatical, in other words, they mainly have grammatical/syntactic functions, whereas forms with more lexical properties contribute mainly to the content, rather than the structure, of the text. Accordingly, the lexical density of a text can be obtained by dividing the total number of words with lexical properties by the total sum of the running words in that same text. Ure (ibid) arrives at a conclusion that the vast majority of the written texts are highly denser than their corresponding spoken ones, that is to say, the lexical density of most written texts is of 40% or even higher, whereas that of the most spoken texts is less than 40%.

With respect to informativity, it has been revealed through a number of studies that the written language tends to incorporate more content words than the conversational language, that is why the former form of language tends to be more informative than the latter (Malvern et al.,2004: 158). What is more, lexical density can work as a measure of informativity. The grammatical meaning encoded via functional/inflectional morphemes is described as schematic and restricted to a closed set, whereas the lexical meaning encoded via content words is fine-grained and related to an open class (Castiglione,2019: 148). It has been pointed out by Schmauder et al. (2000: 1098) that content and functional words tend to be differently perceived in reading, in other words, content words are generally perceived as more communicative and prominent to the extent that they are not likely to be skipped.

An important point that must be taken into account in this respect is that when saying that the written language is denser than the spoken language, this implicitly means that if one starts from the spoken language, the written language tends to appear as more complex. Thus, the idea of lexical density implies the idea of complexity due to informativity development throughout the text (Halliay,1990: 62, & To et al.,2013: 61).

Originally, lexical density has been used mainly in second language studies and it has been used as a measure of the differences between the spoken and the written registers (Malvern et al.,2004: 135). Then, it came to be used in stylistic studies as a characterization of style. It is considered as an obvious stylistic marker in writing. It helps greatly identify the style of a given author, that is to say, when the styles of two (or more than two) authors are compared depending on how many lexical items are incorporated within the given authors' texts, more accurate specifications of style can be obtained. Besides, lexical density can determine how a particular author's style is lexically close to (or far away) form the others (Zaho, 2012: 45-46).

Two important notions are associated with lexical density, namely, information rate and information load. Therefore, highly dense texts are perceived as more difficult than texts with low lexical density (Bosseauk,2004: 114). Nowadays, investigating such aspects of texts is no longer difficult due to the revolutionary technical developments and their role in accessing and processing texts.

It should be noted here that lexical density is only one out of four more ways of measuring style. Mutta (2015: 83) points out that lexical density is included under the umbrella term of lexical richness, which comprises three further additional measures that are: lexical variation (rate of different words), lexical sophistication (difficulty level), and lexical individuality/originality (a particular writer's unique words).

3. Measuring Lexical Density

There have been several measures proposed for gauging the complexity of texts, examples are the type-token ratio, mean word length and lexical density. Lexical density has proved itself as an effective measure of style and in differentiating written and spoken texts (Luis,2004: 207).

Regarding the lexical density measurement, it is essential for the analyst to be able first to differentiate between lexical items (content words) and grammatical items (words with syntactic functions). A very sensitive word in this distinction is the item. This word, item, implies an additional detail concerning lexical forms because an item may differ from a word in terms of length, and it may constitute two parts, for example, call off, take over, stand up, etc, each can work just like a single word. These items are lexical because they are part of the open class (Halliday,1990: 63). Further distinction between the grammatical and lexical items is that of frequency, that is to say, relative to their repeated use, the former forms are more frequent than the latter ones (ibid: 64). One more point concerning lexical density measurement is the question of how to deal with the calculation of the repeated words (Luis,2004: 207).

According to Halliday (1990: 67), the phrase is the scale against which lexical density can be measured, in other words, what is to be measured in a written (or spoken) text "is the average amount of lexical information per clause". He (ibid) presents three simultaneous meanings for this functional unit: the first meaning is that it is "the representation of the phenomena of experience", such phenomena interpretations are culture-dependent, that is to say, they are to be interpreted by the members of the culture, the second meaning is that it is "the expression of speech function" such as the mood categories. However, what is of importance in this respect is the third meaning, which states that the clause is "the bearer of the message, which is organized in the form of theme plus exposition". Other variants have been considered with respect to lexical density. For example, a well-known variant is that of noun density, which is calculated by dividing the total number of nouns per text by the total number of tokens in that same text (Johansson,2008: 65).

Originally, Ure (1971) was the first to propose a measure for lexical density. According to her, lexical density is the ratio of dividing the number of content words by the number of the running words in the text. This measure has been slightly modified later by Halliday (1985) (To et al.,2013: 62). An example will make this point clearer:

The **man reached** his **house** at **night**.

This sentence consists of seven tokens, out of them are four content words and the rest are grammatical forms. In terms of Ure's (1971) original measure, the lexical density of this sentence is obtained by dividing 4/7 and thus the ratio is fifty-five per cent or (0.57).

According to Halliday (1990: 66-67) the clause is a more useful and powerful measure of lexical density, and it helps reflect some insights about language (written and spoken). Accordingly, instead of counting the number of content words to running ones per sentence as Ure does, Halliday has proposed his measure by counting the number of content words per clauses. In his exact words "lexical density will be measured as the number of lexical items per clause". This can be explained via the following example:

The **girl opened** the **window**. The **sun** was **shining** and it was **cool**.

This text consists of three clauses with six content words. Therefore, the lexical density of this sentence is the ratio of dividing 6/3, which is two per cent. In this study, Ure's measure is adopted.

Corpus Data and Methodology

Sampling

Simply, sampling means to divide the collected corpus into smaller sections called samples. From the very beginning of corpus sampling, decisions about the samples length and types are to be made (Kalton,1983: 7).

Regarding the samples length, a single sample is preferred to be of no less than (1,000) tokens minimally and (5,000) tokens or so maximally (Biber,1993: 249). A question can also be considered at this point which is how many samples are to be selected so as to be representative of the corpus being scrutinized. The answer to such question is relative. There are certain factors that can affect the samples number, such as the area of study, the linguistic variation included in that area and the investigation purpose. However, sampling methodologies have been set to meet such requirements. These methodologies have been adopted first in sociolinguistic studies, where the question was what types of population are to be incorporated in a given study so as to be representative of a given population in whole (Nelson,2010: 57, & Meyer,2004: 40).

In corpus linguistics studies, sampling methodology is mainly concerned with the number of samples to be included in the pilot corpus (the corpus under study) so as to adequately represent a particular genre and to which extent they can provide valid generalizations about it (Meyer,2004: 40).

Regarding the samples types, two main types of samples can be adopted: probability and non-probability (Kalton,1983 7).The former type requires careful and prior determination of the data to be investigated, whether these data are people (as in sociolinguistic studies) or texts (as in linguistic or stylistic studies). The reason behind this is to ensure the samples representativeness of the genre or the population under study. The latter type, non-probability sampling, does not depend on such prior considerations, selection or probabilities. Rather, it is based on haphazard or accidental samples (Meyer,2004: 43, see also Kelly,2009: 63-67).

However, probability sampling has certain shortcomings in corpus-based investigations. It requires the use of certain techniques, such as, random, systematic and stratified sampling. Such techniques are criticized by being difficult and time consuming. Consequently, it lacks practicality. Further, since this sampling requires that each element in the sample should be unique in its sake, it follows then that a given person is included only once in a given study. Due to what has been mentioned, linguists tend to prefer non-probability sampling on the expense of the other type (Cole,2015: 29, Kelly,2009: 67, & Hussein,2014: 144). Thus, this paper adopts the non-probability sampling, where each of the three included texts are divided into some number of non-probable samples.

The Corpus of the Study

The corpus of this study consists of electronically stored texts that are available on the internet. No copyright permissions are required since the corpus texts are offered freely by copyright holders. This study is devoted to investigate three well-known novels written by the same author, Virginia Woolf. The three novels are: Mrs.Dalloway(1925), To the Lighthouse (1927) and The Waves (1931).Texts have been downloaded from trusted academic sites, Gutenberg and Library Genesis.

These three novels belong to the same general text type, that is to say, fiction. They have been selected with the aim of investigating their lexical density, and how this can contribute to the richness of the lexical repertoire of style of their author. Further, it has been targeted to determine whether the author's lexical creativity (more specifically, that of content words) has been progressed, kept on a relatively same (or close) level or if even declined.

Corpus Features

There are certain features that characterize the general corpus. When a corpus has only some of these default features, it is considered as special. A number of features have been assumed to characterize the general corpus: (i) quantity: which indicates the corpus size as defined by the number of samples and the number of words (types and tokens) within samples, (ii) quality: which indicates the authenticity and faithfulness in collecting data , that is to say, collecting data from "genuine communications of people doing their normal activities", (iii) representation: which means that data to be incorporated in the corpus should be selected on a broad range of various texts and should reflect as much linguistic features as possible (authentic and balanced), (iv) equality: which means that, as far as possible, samples should be of equal length, (v) simplicity: which means that the text samples should be in a plain and simple format which makes the texts easily accessed by users, (vi) retrievability: this feature is connected with the simplicity of the corpus. Being simple, a corpus should provide other users with a chance to retrieve it and extract whatever material they need to investigate and (vii) verifiability: which means that the data collected are liable to be empirically verified, because a corpus should be open to empirical use; otherwise "its importance is reduced to zero" (Dash & Arulmozi,2018: 17-28).

It should be noted that the previously mentioned features are not fixed or absolute. They allow modification and alternation (ibid). With respect to what has been noted above, the corpus under study has the following features:

1- It consists of three novels, (212,334) tokens, divided into(43) samples and each sample is of maximally (5,000) tokens.

2- Authentic data have been used, that the retrieved texts are the exact copies of their original versions.

3- It is a specialized corpus designed to investigate a specific lexical issue, namely, the lexical density.

4- The corpus three novels are written by the same author and they belong to the same genre, fiction.

5- The corpus is limited to a specific period of time, thus it is a synchronic one.

4.4 Software Tools

Different types of software tools can be used in corpus-based stylistic studies. This study is limited to two software statistical tools, namely, WordSmith Tools and Microsoft Office Excel. These tools are to be sketched respectively in the following sections.

WordSmith Tools (4.0)

WordSmith is a software integrated suite of programs proposed by Scott (1996-). This user-friendly software is used in the automatic lexical analysis processes, reflecting some insights about the way words behave in texts (McCarthy & O'keeffe,2010: 5, & Ho,2011: 144). It consists of six tools, each with a particular function to do in text analysis (Baber & Smith,2005: 123).

These six tools are: wordlist is concerned with words and words clusters generated in terms of frequency or alphabetical order, key words identifies a given text words and compares them with those of another text and concord which investigates the linguistic units in their linguistic context such as collocations (Ho,2011: 144, Bosseaux,2007: 75, & Baber & Smith,2005: 123). The three remaining tools are concerned with other processes such as converting texts, viewing files and dividing the corpus into smaller parts (Baber & Smith: ibid). "Wordlist" is the particular tool adopted in this study.

Microsoft Office Excel

Microsoft Excel, commonly referred to as Excel, is but one part of Microsoft Office suite of programs. It is used to process, calculate and display information in the form of a spreadsheet or worksheet. It has proved itself as a highly efficient tool in processing sizable amounts of data in different areas. In this study, Microsoft Office Excel is used to represent the lexical density ratios and their paths in the form of visual charts.

The Analytical Procedures

Certain steps have been followed in retrieving and processing the data of this study. These steps can be summarized into the following:

1- Texts are retrieved from trusted academic websites, library genesis and Gutenberg. Such sites are quite credible in presenting authentic data.

2- Texts are downloaded in the form of Pdf (or portable document format). Then, they are transformed into txt (or plain text format) to make them accessible via the software program (WordSmith).

3- After clearing them from all additional and study-irrelevant information such as footnotes, websites, numbers and the like, each text is sampled into a number of samples, each sample is of (5.000), or so, tokens.

4- Samples are processed via WordSmith Tools calculating the number and frequency of forms incorporated within each sample.

5- Microsoft Office Excel is utilized to process the outputs of WordSmith Tools by presenting them in the form of graphs.

Data Analysis and Results

Lexical Density Analysis

This section presents the rate and analysis of the lexical density of the corpus to be studied. It is useful to note here that the term lexical density is concerned with the author's use of the lexical forms, more specifically, that of content words. This ratio is obtained by dividing the number of the content words by the total number of tokens. The lexical density analysis of Virginia Woolf's three novels: Mrs.Dalloway, To the Lighthouse and The Waves, is the focus of the following sections.

The Lexical Density Analysis of Mrs.Dalloway

The following table shows the number of tokens and content words within each sample of Mrs.Dalloway, together with the lexical density ratios.

Table (1) The Lexical Density Ratios of (13) samples of Mrs.Dalloway

Samples Number	Tokens	Content Words	Lexical Density
1	5,024	2,465	0.49
2	5,015	2,530	0.50
3	5,030	2,378	0.47
4	5,039	2,470	0.49
5	5,029	2,400	0.47
6	5,028	2,486	0.49
7	5,026	2,596	0.51
8	5,019	2,409	0.47
9	5,018	2,413	0.48
10	5,010	2,432	0.48
11	5,036	2,464	0.48
12	5,021	2,426	0.48
13	3,945	1,789	0.45
Overall	**64,240**	**31,258**	**0.48**

As table (1) shows, the overall tokens number is distributed into (13) samples, each with an approximate number of (5000), or so, tokens. Depending on the numerical data presented in the table above, the following figure gives a visual representation of the general status of lexical density in Mrs.Dalloway.

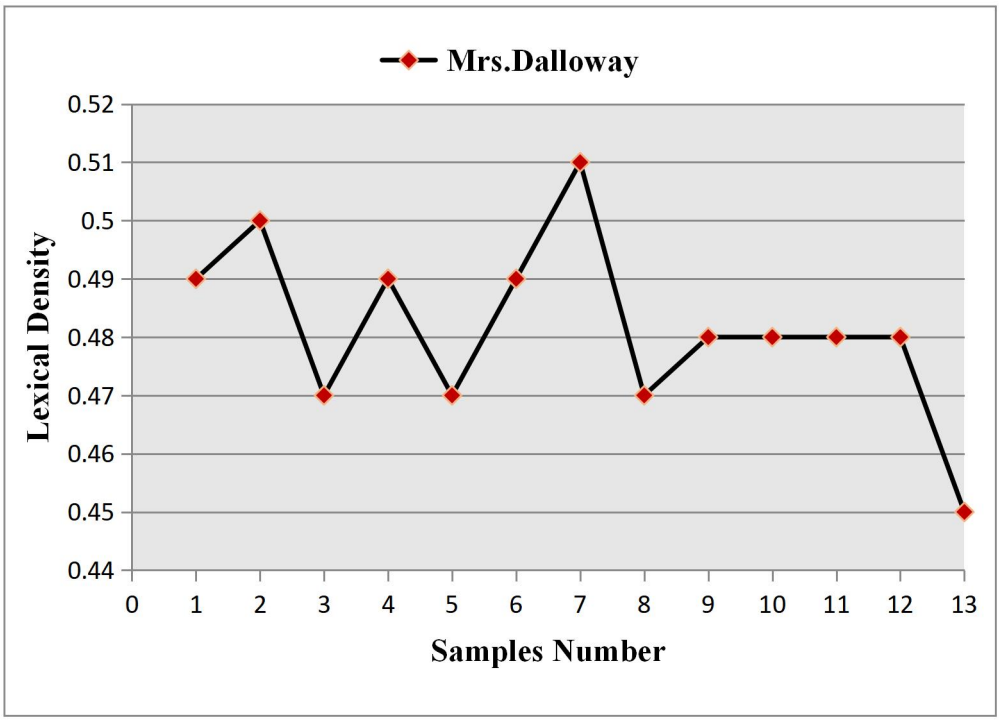

Figure (1) the Lexical Density Ratios of (13) Samples of Mrs.Dalloway

As it can be noticed in figure (1), the ratio of the lexical density takes an inconsistent path throughout the (13) samples. The first two samples score relatively high lexical density ratios, that is to say, (0.49) and (0.50). This ratio drops down to (0.47) within the third sample with only (2,378) content words. It keeps oscillating between (0.49) and (0.47) within the fourth, fifth and sixth samples due to the approximate number of content words captured within these three samples. The author's lexical repertoire reached its climax at sample (7) with (2,596) content words scoring the highest lexical density value within the whole text samples, (0.51). This can give an indication that the novelist condenses her lexical choices at/near the middle of her work putting heavy emphasis on the content words on the expense of the grammatical ones and thus combining the climax of the novel with that of the lexis to, for example, stress an idea or some meaning. From sample (8) up to (12) the lexical density curve comes to take a relatively approximate path, going through very close values that extend between (0.47) and (0.48). The lexical density drops down to its lowest value at the last sample. Due to its low textual size, (3,945) tokens with only (1,789) content words, sample (13) scores a lexical density ratio of only (0.45), which is the lowest value when compared with the previous ones.

The Lexical Density Analysis of To the Lighthouse

The following table presents the statistical data concerning the full textual material of Woolf's To the Lighthouse**.** Its (69,877) tokens are distributed into (14) samples, each with an approximate number of (5000) tokens. The table below displays the number of content words and lexical density ratios as well.

Table (2) the Lexical Density Ratios of (14) Samples of To the Lighthouse

Samples Numbers	Tokens	Content Words	Lexical Density
1	5,027	2,326	0.46
2	5,020	2,346	0.46
3	5,009	2,242	0.44
4	5,014	2,248	0.44

5	5,021	2,234	0.44
6	5,012	2,189	0.43
7	5,010	2,251	0.44
8	5,008	2,214	0.44
9	5,027	2,415	0.48
10	5,026	2,373	0.47
11	5,010	2,320	0.46
12	5,017	2,332	0.46
13	5,017	2,251	0.44
14	4,659	2,078	0.44
Overall	**69,877**	**31,819**	**0.45**

Table (2) displays the statistical outputs obtained for each text sample. The text tokens are distributed into (14) samples, each with relatively the same number of tokens, (5,000). It presents the number of content words together with the lexical density values for both the samples and the whole textual body of the novel. These outputs can be plotted into a visual representation as in figure (2) below.

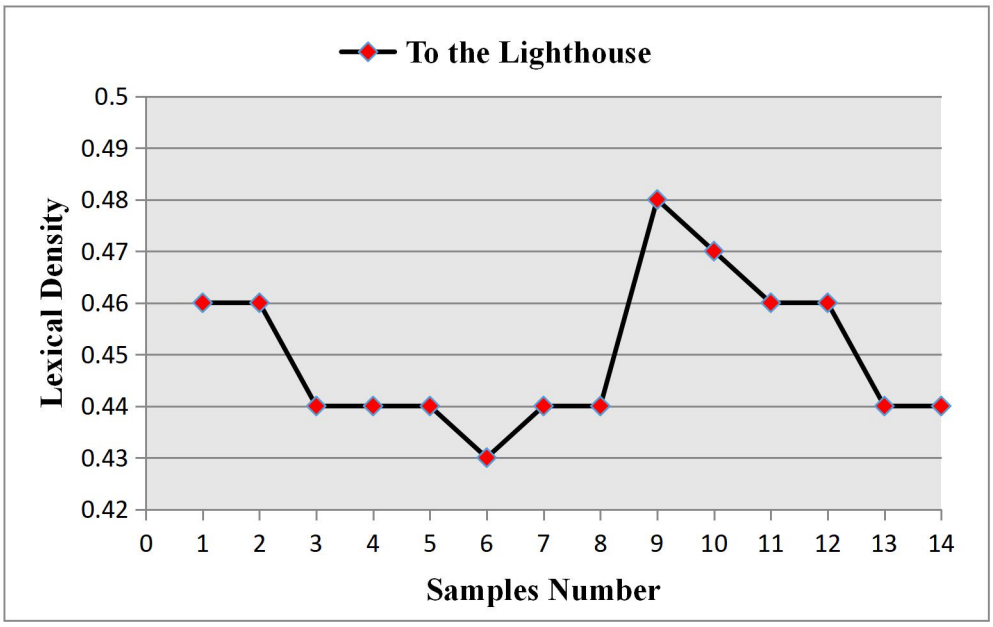

Figure (2) the Lexical Density Ratios of (14) Samples of To the Lighthouse

From the very beginning, it can be noticed that the lexical density curve starts with (0.46) for samples (1) and (2). From sample (3) up to (8), the curve takes an approximated lexical density value that vacillates between (0.44) and (0.43), due to the relatively close number of content words within these samples. The novel reaches its highest lexical density peak at sample (9) with (2,415) content words to score a ratio of (0.48). Then, the ratio starts to slightly drop down to (0.47) within sample (10) and it continues to decrease till the end of the novel. Obviously, the lexical density ratio takes a binary pattern within the last four samples, that is to say, both of sample (11) and sample (12) score the same ratio, (0.46). The same pattern occurs within samples (13) and (14) where their ratio kept at (0.44).

The Lexical Density Analysis of The Waves

The following table displays the statistical details concerning the distribution of the (78,217) tokens of Woolf's The Waves. It reflects the tokens and content words number along with their lexical density values.

Table (3) the Lexical Density Ratios of (16) Samples of The Waves

Samples Number	Tokens	Content Words	Lexical Density
1	5,006	2,558	0.51
2	5,013	2,503	0.49
3	5,016	2,389	0.47
4	5,008	2,437	0.48
5	5,012	2,378	0.47
6	5,009	2,447	0.48
7	5,004	2,326	0.46
8	5,010	2,514	0.50
9	5,008	2,485	0.49
10	5,009	2,496	0.49

11	5,013	2,439	0.48
12	5,020	2,466	0.49
13	5,020	2,480	0.49
14	5,010	2,491	0.49
15	5,021	2,486	0.49
16	3,038	1,458	0.47
Overall	**78,217**	**38,353**	**0.49**

Depending on the information presented in table (3) above, figure (3) below plots these statistical outputs concerning the lexical density of the (16) samples into a visual graph.

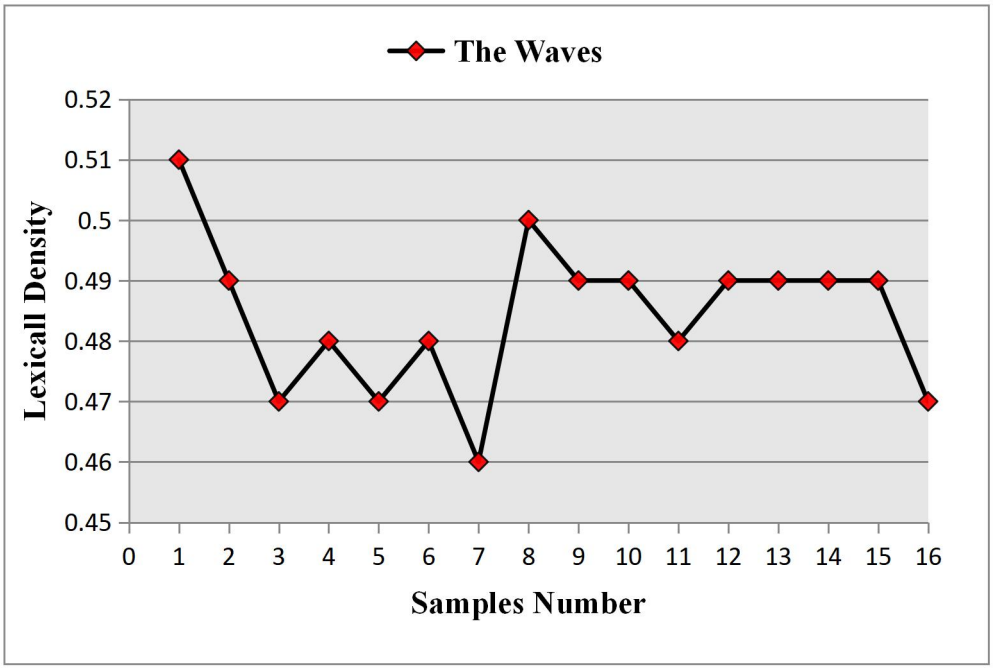

Figure (3) the Lexical Density Ratios of (16) Samples of The Waves

The curve plotted in figure (3) moves in an interesting path, where the higher lexical density value is captured at the very beginning of the novel and reached its lower value near the middle. Sample (1) has the higher content words number (2,558), thus it scores the maximum lexical density ratio, (0.51). Then, though still high, this ratio drops down slowly to (0.49) within sample (2) at (2,503) content words. The curve then takes a zigzagged path within the following four samples. With only (2,389) content words, sample (3) scores a ratio of (0.47). This ratio comes to move up slowly to (0.48) within sample (4) with only (2,437) content words before it comes to drop again to (0.47) within sample (5) at (2,378) content words. Then, the curve again moves up to (0.48) within sample (6). There is a notable shift of value near the middle of the novel, exactly at sample (7). The curve clearly moves down to (0.46) where occurs the minimum number of content words among these (5,000 tokens) samples, (2,326). Then, the curve suddenly moves up to score its second higher value at exactly the middle of the novel scoring a ratio of (0.50) at sample (8) with (2,514) content words. From sample (9) to (15), the lexical density curve takes an approximate path where the ratio vacillates between (0.49) and (0.48) before it finally comes to be stabilized at (0.47) within sample (16).

The Lexical Density Curves of the Three Novels

Regarding the information provided in tables (1), (2) and (3) in the previous sections, the corpus under study is totally divided into (43) samples: (13) samples for Mrs.Dalloway, (14) samples for To the Lighthouse and (16) samples for The Waves. Each sample is of about the same tokens number, (5,000) tokens. What is to be taken into account throughout the following figure is the way the lexical density curves move throughout these three novels. Figure (4) below visualizes the lexical density values accounted for the (43) samples all together.

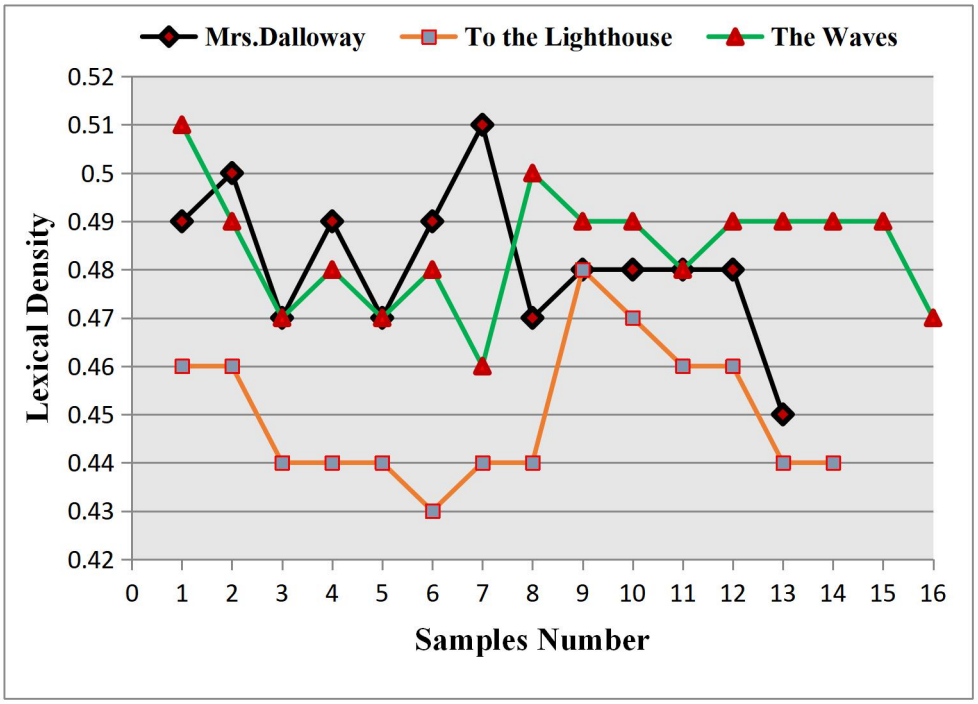

Figure (4): the Lexical Density Curves of (43) Samples of Virginia Woolf's Three Novels: Mrs.Dalloway, To the Lighthouse and The Waves

At first glance, figure (4) above would likely suggest that these three novels were written by three different novelists. The three curves start at different starting points, they take distinct paths throughout the diagram, vary clearly in the positions where the higher and lower lexical density ratios are scored and fall at different finishing points. Nevertheless, they meet and come close to each other at certain points of the path.

To start with the highest starting point which is scored by The Waves, one can see that The Waves scores the majority of the higher lexical density values throughout the diagram. However, its curve approaches and comes in contact with the other two novels curves at some points. The Waves intersects with Mrs.Dalloway near the beginning where both scored the same ratio, (0.47), at samples (3) and (5). It keeps close to Mrs.Dalloway within sample (6) where it holds a lower position with a ratio of (0.48). At sample (7), it seems that the author has exhausted her lexical repertoire with the result that The Waves curve moves downwards to (0.46) and comes close to that of To the Lighthouse, but this does not last long. Soon The Waves curve regains its priority to move straight upwards till it reaches the ratio of (0.50) with sample (8) and keeps on in a distinct path till sample (11) where, again, its curve goes down and comes to meet that of Mrs.Dalloway at a lexical density ratio of (0.48). Then, from sample (12) till the end, The Waves overtakes its competitor where its curve diverges clearly and takes a unique path. It goes on (0.49) through samples (12), (13), (14) and (15) before it finally comes to end its journey and stops at a ratio of (0.47) with the last sample.

The next upper curve of lexical density is occupied by Mrs.Dalloway. It can be noticed that the curve of Mrs.Dalloway occupies a middle position in figure (4) above, where it comes in contact with the curves of the other two novels at different points. Regarding its lexical density, Mrs.Dalloway comes second after The Waves as diagram (4) makes it clear. Tracing the curve of Woolf's Mrs.Dalloway, one can see that it starts at a ratio of (0.49), then it moves upwards to (0.50) within its second sample before it comes to intersect with The Waves at sample (3) sharing it the same ratio, (0.47). The curve moves up to (0.49) at sample (4) exceeding that of The Waves. Then, it moves downwards to again meet The Waves at sample (5) at (0.47). At sample (6), it moves slightly away from The Waves to score (0.49). At about the middle of the novel, exactly at sample (7), Mrs.Dalloway takes its own path away from the other two novels. It moves straightly upwards scoring its maximum lexical density ratio achieving a value of (0.51).

Unfortunately, Mrs.Dalloway did not enjoy this priority for long. Soon, it changes its path to move down to (0.47) within sample (8). With a slight increase of ratio, its curve moves up to (0.48) meeting that of To the Lighthouse at the ninth sample. It moves straightly with this same ratio throughout the three following samples intersecting with the curve of The Waves at the eleventh sample. The curve then takes a downwards path to settle finally at (0.45) at the last sample where it approaches that of To the Lighthouse.

The lowest lexical density curve is that of To the Lighthouse. It occupies the lowest lexical density values as one can notice in figure (4). Its curve moves in a meandered path reflecting vacillating lexical density ratios. Tracing its curve, To the Lighthouse starts with a ratio of only (0.46) at the first two samples. Then, it falls down to (0.44) at the next three ones. At sample (6), it even falls further and moves away from the other two novels to score the lowest lexical density value within the whole corpus with only (0.43). Then, it pulls itself little upwards to reach (0.44) within samples (7) and (8). Suddenly, its curve jumps up to meet that of Mrs.Dalloway with a ratio of (0.48) at the ninth sample scoring its maximum lexical density value. Soon after that, To the Lighthouse starts its path downwards again to score (0.47) at the tenth sample. This decline continues within the eleventh and twelfth samples where both score (0.46). This ratio decreases further towards the end where it reaches (0.44) at the thirteenth and fourteenth samples closing up the novel lexical density curve.

The Results

Finishing up with the quantitative analysis, it is the time now to interpret the statistical information presented in the previous figures in a qualitative manner. To attempt this, the lexical density values plotted for the corpus under study are all brought together in a separate figure. This helps greatly in presenting a rather precise visual representation of the three novels' curves, with the result that it would be easier to detect the rate of the novelist's lexical density development/decline throughout these three novels.

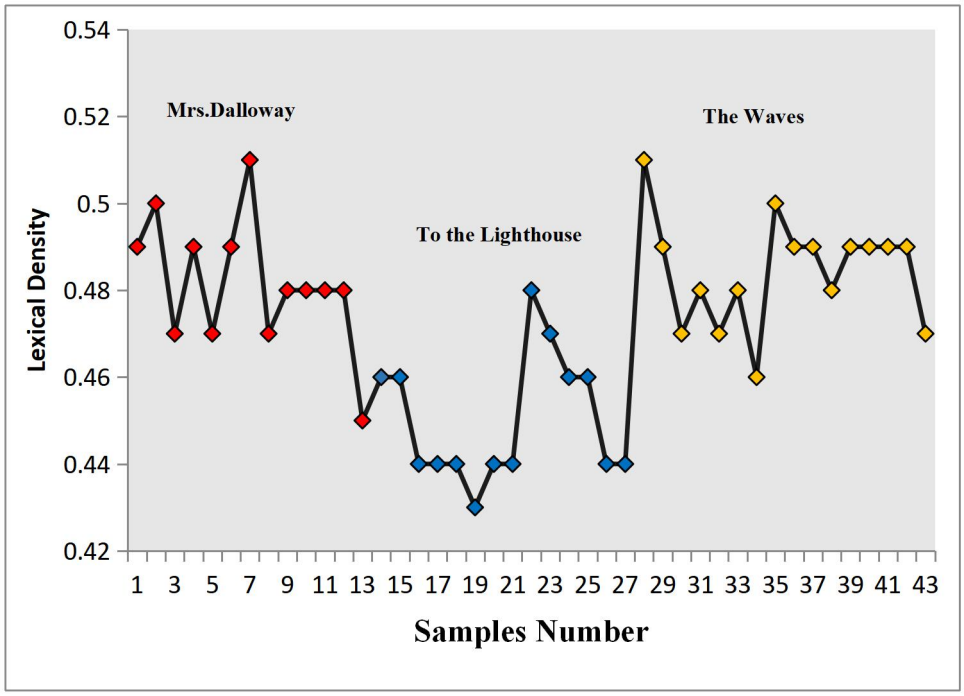

Figure (5): the Order of the Lexical Density Curve for the Three Novels: Mrs.Dalloway, To the Lighthouse and The Waves

Following the curve plotted in figure (5) above, one can notice that the highest two lexical density peaks are equally occupied by Mrs.Dalloway and The Waves giving the impression of how lexically rich they are. Nevertheless, this does not mean that they have the same level of lexical density along the curve. A close look at the diagram will show that The Waves occupies the majority of the higher lexical peaks due to its higher lexical variety. It incorporates (38,353) content words out of (78,217) tokens with the result that it proved itself as the richest amongst others.

Actually, The Waves can work as an idealistic representative example of Virginia Woolf's writing style, to the extent that it is nowadays considered 'as her masterpiece' which is 'most demanding and entirely unique in form' and one of the most important twentieth century narrative achievements (Balossi,2014: 6 & Rollyson,2001: 1030-1031). Due to its high lexical variety and its author's "highly artificial trick in writing", it is characterized by most critics as 'extremely difficult'. This complexity and difficulty have been confessed also by many critics and scholars. While

some are attracted by its high stature, some others regard it as Woolf's most experimental work as they are taken by its formal innovation (Balossi: ibid, Goldman,2006: 69, & Marsh,1998: 38).

Mrs.Dalloway, regarding its lexical density (more precisely, (31,258) content words out of (64,240) tokens), comes to occupy the second position after The Waves. With the result that the last position is left to be automatically occupied by To the Lighthouse scoring almost nearly the lowest lexical peaks (as figure (5) above makes it clear) with only (31,819) content words out of (69,877) tokens. The structure of To the Lighthouse has been characterized as being 'simple and straightforward'(Balossi,2014: 59).

Tracing up the path of the curve in figure (5) chronologically, one can notice that the path of the three novels, so to speak, takes a curved shape. While the novelist's lexical repertoire is flourishing in Mrs.Dalloawy, it comes to suffer a decline in To the Lighthouse before it moves upwards again to prosper and progress greatly in The Waves.

To sum up, it is now possible to determine how, where and when the author's style lexical density does really and clearly reflect itself. Therefore, The Waves reflects a remarkable progress in the novelist's style and her lexical resources throughout its high lexical density ratio, as clearly reflected in its lexical density curve. Thus, it overtakes its two other companion texts moving up to occupy the priority position with no rivals.

Conclusions

Closing off this study, it should be stressed that the revolutionary technological development, more specifically that of computers, adds much to the field of corpus linguistics which in turn contributes greatly to corpus stylistics. Thus, corpus stylistics borrowed its tools from corpus linguistics. Making use of such tools, corpus stylistics started to flourish as a fairly qualified field in dealing with large amounts of electronically stored texts, a matter which is extremely difficult and time consuming if it were to be handled manually. It comes to have an exceptional and unprecedented ability to handle sizable amounts of data (just like the study being tackled which consists of (212,334) tokens) in a relatively feasible way saving much effort and time.

However, the lexical density curves of this study have shown notable results concerning the novelist's style lexical density throughout these three novels. The three novels curves and their statistical analysis outputs show how the author's lexical density moves within each novel and how distinct paths their curves tend to take; though they are produced by the same writer. So that, while the writer's lexical density reaches its optimum level in The Waves, it reflects an opposite situation in To the Lighthouse. The lexical density ratio proved its power in providing insightful knowledge and strict statistical accounts concerning the author's lexical repertoire development and decline.

Throughout the passage of time, writers can develop their lexical creativity and, thus, their style and lexical repertoire become richer. The opposite situation is equally expected, that is to say, through the course of time, writers may suffer a decline in their writing creativity and thus their lexical creativity decreases. This is a situation that can be observed in Woolf's three novels. Where she presented a rich text within Mrs.Dalloway, such lexical richness seems to be exhausted within To the Lighthouse, then her lexical repertoire comes to step forward again within The Waves overtaking the other two novels. This gives the impression that the novelist's lexical repertoire has refreshed and flourished once again during that period of time.

REFERENCES

Baber, E. and Smith, D. (2005). Teaching English with Information Technology. London: Modern English Publishing. Retrieved from https://books.google.iq/ (19 Aug.2019). P.123.

Baker, P, Hardie, A, and McEnery , T. (2006). A Glossary of Corpus Linguistics Edinburgh: Edinburgh University Press. PP.47-76.

Balossi, G. (2014). A Corpus Linguistic Approach to Literary Language and Characterization: Virginia Woolf's The Waves. Amsterdam: John Benjamins Publishing Company. PP.6-59.

Biber, D. (1993). "Representativeness in corpus design". Literary and Linguistic Computing. Vol.8(4). PP.243-257.

Bosseauk, C. (2004). Translation and Narration: A Corpus-Based Study of French Translations of Two Novels by Virginia Woolf. United States: ProQuest LLC. P.114.

Bosseaux, C. (2007). How does it Feel?: Point of View in Translation: The Case of Virginia Woolf into French. Amsterdam: Rodopi. P.75.

Castiglione, D. (2019). Difficulty in Poetry: A Stylistic Model. Switzerland: Palgrave Macmillan. P.148.

Cole, A. (2015). The Implications of Consumer Behavior for Marketing: A Case Study of Social Class at Sainsbury. Hamburg: Anchor Academic Publishing. P.29.

Dash, N. and Arulmozi, S. (2018). History, Features, and Typology of Language Corpora. Singapore: Springer. PP.17-28.

Goldman, J. (2006). The Cambridge Introduction to Virginia Woolf. Cambridge: CUP. P.69.

Halliday, M. (1990). Spoken and Written Language. Oxford: OUP. PP.61-67.

Ho, Y. (2011). Corpus Stylistics in Principles and Practice: A Stylistic Exploration of John Fowles' The Magus. London: Continuum. P.144.

Hussein, K. (2014). A Computational Linguistic Perspective of Authorship and Plagiarism. German: Lambert Academic Publishing. P.144.

Johansson, V. (2008). "Lexical diversity and lexical density in speech and writing: a developmental perspective". Lund University Working Papers, Vol.53. PP.61-79.

Kalton, G. (1983). Introduction to Survey Sampling. London: SAGA Publications. P.7.

Kelly, D. (2009). "Methods for evaluating interactive information retrieval systems with users". Foundations and Trends in Information Retrieval. Vol.3, Nos,1-2. PP.1-224.

Kondal, B. (2015). "Effects of lexical density and lexical variety in language performance and proficiency". International Journal of IT, Engineering and Applied Sciences Research (IJIEASR), Vol.4, No.10. PP.25-29.

Luis, M. (2004). " Lexical density and noun group elaboration as indicators of written style in native and EFL compositions". In, Anderson, J., Cabanas, J. and Zapata, J. (eds.)2004. Linguistic Perspectives from the Classroom: Language Teaching in a Multicultural Europe. Universide de Santiago de Compostela. Retrieved from https://books.google.iq/ (17 Aug.2019). P.207.

Malvern, D., Richards, B., Chipere, N. and Duran, P. (2004). Lexical Diversity and Language Development: Quantification and Assessment. Hampshire: Palgrave Macmillan. PP.135-158.

Marsh, N. (1998). Virginia Woolf: The Novels. New York: Macmillan. P.38.

McCarthy, M. and O'keffee, A. (2010). "Historical perspective: what are corpora and how have they evolved?". In, O'keffee, A. and McCarthy, M. (eds.)2010. The Routledge Handbook of Corpus Linguistics. London: Routledge. P.5.

Meyer, C. (2004). English Corpus Linguistics: An Introduction. Cambridge: CUP. PP.40-43.

Mutta, M. (2015). "Lexical richness in expository essays written by learners of L3 French". In, Pietila, P., Doro, K. and Pipalova, R. (eds.) 2015. Lexical Issues in L2 Writing. Newcastle: Cambridge Scholar Publishing. Retrieved from https://books.google.iq/ (15 Aug.2019). P.83.

Nelson, M. (2010)."Building a written corpus: what are the basics?". In, O'keeffe, A. and McCarthy, M.(eds.)2010. The Routledge Handbook of Corpus Linguistics. London: Routledge. P.57.

Rollyson, C.(ed.) (2001). Notable British Novelists. Vol.3. California: Salem Press, Inc. PP.1030-1031.

Schmauder, A., Morris, R., and Poynor, D. (2000). "Lexical processing and text integration of function and content words: Evidence from priming and eye fixations". Memory and Cognition, 28(7). P.1098.

To, V., Fan, S. and Thomas, D. (2013). "Lexical density and readability: a case study of English textbooks". International Journal of Language, Culture and Society. PP.61-71.

Ure, J. (1971). “Lexical density and register differentiation”. In, Perren, G. and Trim, J. (eds.) 1971. Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge 1969. Cambridge: CUP. P.445.

Yule, G. (2010). The Study of Language. Cambridge: CUP. P.68.

Zaho, M. (2012). “The art of balance: a corpus-assisted stylistic analysis of Woolfian parallelism in To the Lighthouse”. International Journal of English Studies (IJES), Vol. 12(2). PP.39-58.