

Corpus Stylistic Analysis of Thomas Harris' *The Silence of the Lambs*

Eman Adil Jaafar

University of Baghdad, Baghdad, Iraq

1. Introduction

First of all, the Latin word *corpus* (plural "corpora") refers to a collection of texts. It has the meaning of "body". The texts are stored in an electronic database. Baker, Hardie & MacEnery maintain that "Although a corpus does not contain new information about language by using software packages which process data, we can obtain a new perspective on the familiar"(48-49).

Corpus stylistics is a branch of computational linguistics as Wales (1989) points out. It is developed in the late 1960s. It helps to investigate certain characteristics of the data like words length and sentences based on statistical and computer-aided tools to study a number of issues related to style (85).

One important issue needs to be given attention to here in the study, McIntyre (2015) sheds the light on the difference between corpus stylistics and corpus linguistics. He asserts that many works have developed the sub-branch of stylistics namely, corpus stylistics, but none of them give an accurate definition of corpus stylistics and differentiate it from corpus linguistics (60).

Then, he defines corpus stylistics as "corpus stylistics is simply corpus linguistics with a different object of study (literature as opposed to non-literary language)". Besides, he demonstrates that the difference between them is that corpus stylistics is not only borrowing tools from corpus linguistics but it makes itself unique by using qualitative tools and techniques of stylistics to analyze texts with the help of computational methods (McIntyre 60).

Semino and Short (2004) assert that doing a corpus-based work is highly quantitative in nature, but this matter of fact does not mean that qualitative analysis should be excluded. On the contrary, if both are combined they help to reach to a higher degree of understanding of literary works or any other selected data (7).

Jabbar (2014) cited in (Jeffries & McIntyre 181-185) explains that there is a great significance of using corpus-based analysis. It assists to deal with large texts with relatively short time. In addition, it can achieve the objectivity that stylisticians seek for. It can reveal crucial features that can be missed in the manual analysis (23).

Mahlberg (2014) draws attention to the fact that corpus stylistics links principles from corpus linguistics and literary stylistics. Moreover, following a corpus stylistic approach can help to locate the linguistic innovative uses of the language of texts by quantitative methods (378-380).

Mahlberg (2014) quoted Carter (2010) who describes the approach of analyzing literary texts and asserts that this method is conducted in a relatively objective way and for the optimal achievement, it should be exerted by a relatively subjective way (67).

2. Corpus Linguistics and Stylistics: a Linkage between Quantitative and Qualitative Study

The majority of stylistic studies including articles, books and other works use qualitative methodologies in order to analyze literary texts. Researchers can refer to particular features that are related to meaning. In my previous works (cf. Jaafar 2014, 2016), I followed qualitative methodologies to reach literary texts. One of these articles concentrates on Leech and Short's (1981) methodology that focuses on examining aspects like grammar, adverbs, adjectives, figures of speech and other stylistic aspects. The main aim was to find a rigorous interpretation of the text by scrutinizing the most prominent and stylistically deviant features by means of manual analysis (75-79).

In regard to what is mentioned, the current study follows a quantitative approach along with a qualitative one. Quantitative approach is known as a counting approach which allows counting words, items, certain grammatical aspects, for instance, the use of (ed) past tense. Due to the recent advances in using technology in research, stylistics specifically, corpus stylistics, benefits from this phenomena to conduct quantitative studies to examine literary texts.

Historically, as Fialho & Sonia (2014) explain that there are two branches of quantitative methods related to stylistics. First, stylostistics (statistical stylistics, or stylometrics) dates back to the mid of the 1800s that has the basic interest of studying frequencies or repeated patterns of texts. The final results can help to

show whether the repeated words or aspects happened to occur as a result of 'chance or choice' by the author. The origin of stylometrics is Saussurean linguistics also Charles Barley's 1909 study (330-332).

Second, quantitative studies in corpus stylistics have different origins or backgrounds. Actually, the roots belong to Russian Formalists especially the Moscow Linguistic Circle of the 1920s. Fialho & Sonia clarify the main aim of these studies was " how textual patterning worked in literary-aesthetic communication. In this sense, they go beyond linguistic analysis by combining other fields such as psychology and sociology." (330-332).

3. Methodology

This study follows Mahlberg & McIntyre's (2012) methodology. This model focuses on one single literary text by one author. They explain that studying one text may be considered as a "small sample of data" but then they assert that this text is still regarded as part of a corpus (206).

In their (2012) article, they show how a word like "table" refers to the fictional world of the novel; *Casino Royale* since this word has a tangible meaning and it has a reference to one of the concrete objects that can be found in a place like a casino. Furthermore, the word "gambler" indicates to the theme of the novel. In other words, it has the main function of being a 'thematic signal' keyword. It is worth mentioning that, not all the words have these straightforward indications. This is basically because words can have polysemous (the coexistence of many possible meanings for a word) meanings (209).

Mainly, this work aims at examining **keywords**, **key semantic domains** as well as **clusters**. Firstly, keywords can be defined as the most frequent or repeated words in a single text or group of texts in comparison to a reference corpus. Words are the crucial part of any corpus study. There are three groups of words in general: proper nouns, content and function words. Mahlberg & McIntyre point out that the most common words are function words. They work as the constituents of any text. However, content words are the carriers of meaning and writers' messages. For this reason, they are important for studying (384).

Secondly, Gliozzo & Strapparava define semantic domains as "fields characterized by lexically coherent words. The lexical coherence assumption can be exploited for computational purposes because it allows us to define automatic acquisition"(5).

Both keywords and key semantic domains are used to describe the fictional world, setting as well as the thematic signals focusing on the characters of the text. This is done by using Wmatrix3 software package (Rayson, 2008).

Thirdly, clusters will be checked with reference to Mahlberg's (2013) book which is mainly about this particular area of study. AntConc (Anthony, 2011) is the tool for achieving this goal. In this paper, the researcher will use this model as a basis for the study. The following steps are to be followed:

- 1- Examining keywords of the novel *The Silent of the Lambs*. Keywords can guide readers to the meaning of the text. Keywords are considered the basis of the world text or the fictional world.
- 2- Finding the connection between the resulted keyword and key semantic domain analysis.
- 3- Looking for lengthy n-grams, and identifying their significance, value and their connections to the overall meaning.

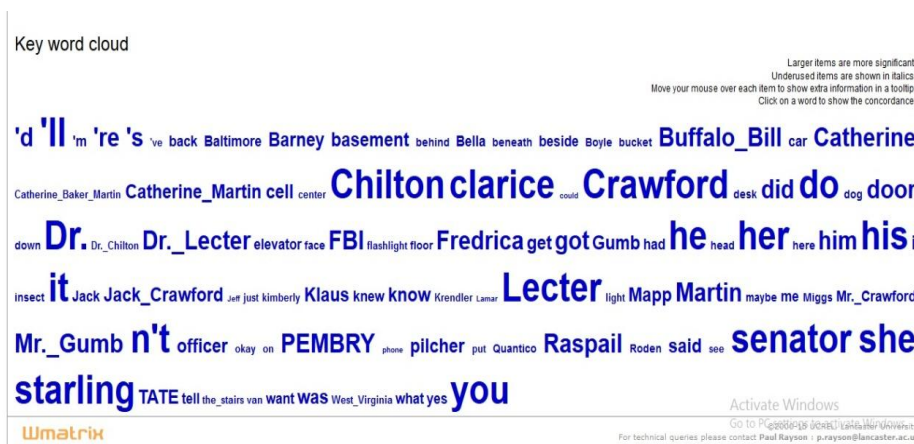
3.1.What is Wmatrix3? How does it Work?

A software tool works as an analysis system to analyze English texts. Paul Rayson at Lancaster University comes up with this tool during his research project. Wmatrix3 works as a concordance tool that can calculate keywords, a key part of speech tags and key semantic categories in a text. (Paul Baker, Andrew Hardie and Tony McEnery 168). It also provides a web interface to CLAWS(the Constituent Likelihood Automatic Word-tagging System) part-of-speech tagger and USAS (The UCREL semantic analysis system) semantic tagger (corpus connotation tools). In addition, methodologies related to corpus linguistics for example frequency lists and concordances ("*WmatrixCorpus Analysis and Comparison Tool*" par.1-3). One thing needs to put into consideration, studying n-grams or clusters in this tool is not activated yet. Therefore, AntConc in this paper will fill this gap.

Explaining how Wmatrix3 works might not be the main aim of this study. However, it is necessary for new users or researchers who are interested in using this tool to conduct a corpus stylistic study to know its major basics.

The tool has both simple and advanced interfaces. The simple one consists of four functions. These include finding a list of words and their frequencies, searching for a particular word, word clouds which allow seeing the significance of the words which are illustrated by the font size in the tool. The data is compared to different corpora of the BNC (e.g. BNC Written Sampler, BNC Spoken Sampler, BNC sampler CG Business, etc). The word clouds are represented in different sizes of

fonts large and small fonts. The larger items are more important and have more significance. The below image illustrates the fact:



Screenshot (1) Key Word Clouds by Wmatrix3

From screenshot (1), the word clouds appear alphabetically. It is quite obvious that the most crucial items or words are the proper names of the characters in the novel who are the main presenters of the events of the texts.

While the advanced interface has other different tags. For example, the semantic and part of speech tags.

[Tagging > Tag Wizard... | My Tag Wizard... | Domain Tag Wizard... | Load file without tagging...]
 [Folders > My folders | Details | Create... | Delete... | Archive... | Extract... | Join... | Share... | NEW! CrossTab... | Empty TRASH]
 [Options > Switch to Simple Interface | Edit user options...]
 [Help > Contents | Availability | Tagsets: POS & Semantic | USAS: Lexicon & MWEs & Context rules | Updates | Feedback]

[You are here > My folders > file20083_1466717276 | File details | Summary sheet]

	Frequency list	Concordance	N- & C-grams	Collocation	Keyness analysis
Word	Word only (Sorted by: Frequency, Word)	Word		Word	Key words compared to: BNC Sampler Spoken <input type="text"/> <input type="button" value="Go"/>
Part of speech	POS only (Sorted by: Frequency, POS) Word and POS (Sorted by: Frequency, Word, POS)	POS			Key POS compared to: BNC Sampler Spoken <input type="text"/> <input type="button" value="Go"/>
Semantic	USAS Tag only (Sorted by: Frequency, USAS tag) Word and USAS tag (Sorted by: Frequency, Word, USAS tag)	USAS tag		Word-Semantic	Key concepts compared to: BNC Sampler Spoken <input type="text"/> <input type="button" value="Go"/>

Screenshot (2) The Advanced Interface of Wmatrix3 Tool

3.2. AntConc

AntConc is a free software toolkit for extracting keywords, concordances as well as n-grams designed by Laurence Anthony(2011). In this study, AntConc helps to identify and calculate long clusters that facilitate the qualitative analysis, which in turn enables to a better comprehension of the chosen text.

4. The Data

The Silence of the Lambs is an American novel by Thomas Harris published in 1988. It was written as a sequel to Harris's 1981 novel *Red Dragon*. Basically, the novel depicts investigations of heinous crimes, which are committed by a serial killer, James Gumb, whose nickname is Buffalo Bill. The criminal used to skin his women victims to create a woman's suit for himself. The heroine of the novel, Clarice Starling, is a student at the FBI Academy. She wants to find clues about the killer by asking the cannibal, Dr. Hannibal Lecter, while he is imprisoned for killing nine people and eating them. Both Lecter and Gumb were in a psychiatric hospital. That is why Clarice wants Lecter's help to solve this case (Huntley, par.1-5).

The reasons for choosing this particular novel *The Silence of the Lambs* include; firstly, the novel gained a noticeable appeal from the audience and it got several awards, being a horror fiction text attracts readers heavily. Also, the text is available in electronic form freely without any copyright restrictions.

5. Thematic Signals or Indicators

There are certain words used to indicate in a direct or indirect way the main theme of the text, these are called 'thematic signals'. These words have importance over others. They are simply the carriers of meaning or the clues to the hidden codes in the text.

Roughly speaking, the novel has the following thematic events; horror, thriller and psychological themes. The novel comprises 95,145 words.

The word 'dead' for example, occurs (45) times throughout the text. It leads to the basic theme in the novel. As mentioned before, the story is about the inspection of crimes committed by a serial murderer. Similarly, the word 'victim' with (22) occurrences has a noticeable thematic value. The word 'body' occurs (60) times,

indicating also to the theme of murdering in the novel. The occurrences or the concordances of this word in the text lead the reader to this interpretation. Here are examples of the associations of the word within the context:

Screenshot (3) Thirty Three Examples of Successive Concordances of the Word 'body'

60 occurrences.		Ext
ance in Baltimore . On March 25 his	body was discovered seated in a pew in a	1
. Turned at an owlish angle to the	body beneath , it gaped stupidly at Starl	2
ve . Will you tell me if there 's a	body in there ? Camera 's off , just betw	3
ck hunters in West Virginia found a	body in the Elk River around daylight . I	4
he did the rest .) He dumped each	body in running water when he was through	5
. " We 're lucky on this one . The	body was snagged on a trotline--- a fishi	6
word assume . " He 'd drop off the	body first , would n't he , in case he go	7
it down to zip if he did n't have a	body in his car . So how about drawing ve	8
that a death is questionable , the	body is sent on to Claxton Regional Medic	9
She turned around and looked at the	body on the table . " They should have pu	10
posures , Starling photographed the	body . The victim was a heavy-hipped youn	11
en in it more than a few days . The	body was flayed neatly from a clean line	12
se his lips as he walked around the	body . Starling peeled off her glove to t	13
oks that had entangled and held the	body in the moving river was still wrappe	14
Give me the flashlight . " " When a	body comes out of the water , a lot of ti	15
. She looked at Crawford across the	body . He nodded . It only took her a sec	16
hat , " Lamar said . Facedown , the	body was easy to fingerprint . Starling h	17
er case . It was different with the	body and problem out of sight . In this s	18
ggage compartment . " He hauled the	body in something big enough for the body	19
ody in something big enough for the	body to be stretched out on its back , "	20
do n't know how it got there . Her	body was in the Elk River in West Virgini	21
s were bound so tightly against the	body , they might have been carved in low	22
othing that would normally infest a	body outdoors , and it would n't be in th	23
span of nearly six inches , a furry	body , and slender antennae . 'Erebus odo	24
is features , but she-could see his	body plainly . He had on pressed khaki tr	25
ch afraid Catherine Baker Martin 's	body had been found . When Buffalo Bill f	26
ight have killed her and dumped her	body with a bug in the throat . Maybe Cra	27
It reached eighteen days before the	body floated . " So Catherine Baker Marti	28
t Virginia yesterday . They found a	body over there , Buffalo Bill 's latest	29
oohed it . " " And after he saw the	body in West Virginia ? " He talked to hi	30
ter bounded to his feet , his tithe	body suddenly grotesque , bent in a gnomi	31
om up and she had some sense of her	body 's position . Her shoulder , hip , a	32
on can stand to look at a mutilated	body , " Crawford said , crumpling his cu	33



From the above screenshot, it is clear from the right context that the words “water” and “river” are associated with the location of the dead bodies. This repeated co-occurrence of words has an important significance. In line (4), Elk River is mentioned to refer to the location where some duck hunters find one of the victims' bodies. In lines (5) and (14) there are two textual phrases: “running water” as well as “in the moving river.” Line (15) gives the indication of taking the body “out of

the water.” Again, line (21) gives information about a body, which is found in the Elk River. In line (28), the word “floated” pinpoints to water. From the resulted concordances, the computer facilitates locating such information: It provides readers with fast access to information based on textual evidences.

Table (1) Fictional world and thematic signal keywords with examples of subgroup based on Mahlberg and McIntyre's (2012, 210).

Category	Example keywords
<p>Fictional world</p> <p>Characters:</p> <p>Names</p> <p>Body parts</p> <p>Clothes and accessories</p> <p>Settings and props:</p> <p>Places</p>	<p>Starling, Lecter, Gumb, Buffalo Bill, FBI</p> <p>Skin, finger, mouth</p> <p>shirt, collar, clothing, dress, cuff, bags, hat, coat</p> <p>basement, hospital, Elk river</p>
Thematic signals	Murder, skin, victim, dead, body

6. Keywords in *The Silence of the Lambs*

Table (2) below contains the top 15 keywords produced by the Wmatrix3 tool comparing the selected novel with BNC Sampler Written. In addition, the table illustrates the frequency of keywords in the novel. The language of the text contains words that are grouped into three types or categories:

- a. Words that can give ideas about the story or the theme of the novel.
- b. Proper names can introduce readers to the dominant characters in the text.
- c. Grammatical words which can indicate something useful about the style of the writer concerning the novel.

Table (2) The Top 15 keywords of *The Silence of the Lambs* Compared to BNC Sampler Written.

keyword	Freq in the novel	LL value
starling	682	3318.82
she	1272	1762.84
her	1193	1702.78
you	1577	1639.08
he	1672	1460.48
Crawford	270	1313.90
Lester	255	1240.91
Dr.	209	862.60
n't	727	856.
clarice	125	608.29
chilton	109	530.43
senator	114	514.19
it	1536	488.15
'll	277	479.41
his	913	478.54

Apparently, Table (2) shows that words such as (she, you, 'll, it, his, etc) are dominant in terms of frequency. Nevertheless, this does not mean that researchers neglect a word like 'body', which is illustrated previously. Computational retrieval of keyness should be checked and interpreted by manual examination to get a reasonable and a subtle way of reading the text by means of corpus stylistics study.

It is worth mentioning that Mahlberg and McIntyre maintain that the analysis of the concordance lines is necessary to point out keywords related to the "thematic signals" and "fictional world" in order to search for the meaning within the context (209).

Keywords that are related to the fictional world are usually concrete and they indicate to:

- i. Characters: This can have subcategories like names of persons, body parts, clothes and accessories, etc.

- ii. Setting: under which also certain subcategories comprise places, equipments, etc.

The researcher endeavors to identify which character plays the most important role in the events of the novel. The results show the following:

- Starling (Clarice's family name) appeared 682 times.
- Lecter is mentioned throughout the novel 255 times.
- Gumb (James' family name) is repeated 40 times, while his nickname Bill is mentioned 14 times only.

From the above results, Starling proved to be the most important character in the event of the novel since her role is to investigate the crimes that are committed by James Gumb. Even the title of the novel is related to the leading character. It is about the dream she used to see during her childhood.

Keywords of the fictional world are tangible. That is why they are considered as text-centered. While thematic signals are reader-centered, they tend to be somehow subjective. They depend highly on the readers' intuitions.

7. Key Semantic Domains

Before giving examples of key semantic domains from the text, it is necessary to explain the concept of semantic domains and its origin. According to Gliozzo & Strapparava (2009) semantic domains are originally based on the concept of semantic fields that is proposed by (qtd .Trier, 1931). This concept has the limitation of not being objective concerning the differentiation among various semantic fields. The main idea of Trier's theory is words are related to other words. They are arranged in clusters (sets of words) and have the same ideas or meaning. These are called **semantic field**, and the study is called **semantic field theory** which is based on Saussure's structural semantics (18). .

Gliozzo & Strapparava shed light on how to find a solution to Trier's theory limitation. Semantic Domain is the solution resulted from the structuralist approach in semantics with Ludwig Wittgenstein's principle of language game (meaning- in - use). This principle simply means that words are meaningful when they occur in an identified situation or context. In short, readers can discern or locate the meaning of words for concrete contexts (18).

In fact, semantic domains have two advantages. The first is to point out which domain label best fit to words (lexis). For instance, it helps to locate these sets of

domain words; flower, tree and photosynthesis to the suitable or relevant domain, which is botany. The second advantage is a broad one it is useful to "text categorization" (Gliozzo & Strapparava 21).

Table (3)Key Semantic Domains from the novel

Semantic domains		Examples
.1	Anatomy and physiology	hair, hands, eyes, arm, feet, chest, palm, pens, cell, cheeks, jaw, tongue, finger[...]
.2	Medicine and mental treatment	Dr., hospital, tablets, Alka-seltzer, psychiatrist, mental hospital, nurse, prescribed[...]
.3	Discourse bin	all right, no, good morning, hello, yes [...]
.4	Knowledgeable	Knew, recall, remember, profiling, information [...]
.5	Light	Shine, light, beam, daylight, rays, strobe, gleam, splendor, firelight
.6	Living creatures, animals, birds	Starling, beagle, snake, bird nest, monster, tail, chicken, lamb, pecking, swans, tiger, bear, print, wing, frog, mouse. Caterpillar[...]
.7	Clothes and personal belongings	Windbreaker, shirt, collar, clothing, dress, cuff, bags, hat, coat, umbrella, garment[...]
.8	Sensory sight	Reflection, see, look at, look him over, watched, glimpse, peeped, stared, sight, spot, witness, visible, peek
.9	Part of building	Floor, doors, suite, moat, wall, lobby, corridor, elevator, stairs[...]
.10	Object generally	Plate, file, things, knife, bolt[...]
.11	Direction and location	Ground, outer, standing, top, center, here, leaned, ahead, around, somewhere, behind, there, middle, end [...]

In allusion to the above-mentioned information, the retrieval results by using Wmatrix3 show that the occurrences of words of the semantic domain (Anatomy

and psychology) reach to (1910) times in the novel. This number highlights the importance of words within this domain and their connection to the main theme of the text.

In a similar way, the occurrences of words of the semantic domain (Medicine and mental treatment) reach to (507) times. This reflects the same fact mentioned previously.

8. Clusters or n-grams

Clusters (n-grams or lexical bundles) can be defined as repeated sequences of words. According to this study, short clusters, i.e. 2-grams, 3-grams and 4-grams cannot provide the required data to support the qualitative interpretation. Therefore, this research focuses on long sequences of words only. To put it differently, the retrieval of 8-grams, 7-gram, 6-grams and 5-grams supports the study. Malhberg (2013, 45) pinpoints that when the number of n-grams is lengthy, their frequency usually goes down. The following table (4) demonstrates the fact:

Table (4) the number of n-grams in *The Silence of the Lambs* occurring at least 3 times

8-grams	7-grams	6-grams	5-grams	4-grams
1	3	9	51	391

8 &7-grams

In this part of the analysis, 8-grams are investigated. It is worth noting that 8-grams is the maximum number of grams in the selected novel. Also, the maximum number of frequencies is 3 times. 4 times frequency is not found. The tool AntConc reveals only one 8-grams that are illustrated in the table below:

Table (5) 8-grams in *The Silence of the Lambs*

Rank	Frequency	8-grams
1	3	the Baltimore state hospital for the criminally insane(ch.2 , 21, 31)

The most repeated 8-grams "the Baltimore State Hospital for the criminally insane" is the name of the hospital where Dr. Lecter and Gumb used to be captive in. It is an adverb of place with the description of the specialty of the hospital for those who are criminals and having psychological problems. From this analysis, it is

apparent that the 8-grams are associated with important characters in the novel who participate much to the events of the novel.

Following the same pattern, 7-grams results show the same previous example with only 3 times frequencies without the definite article 'the' and the other case drops the noun 'insane'.

"Baltimore State Hospital for the criminally insane".

"the Baltimore State Hospital for the criminally".

6-grams

These six-word clusters "the Baltimore State Hospital for the criminally insane" can also show repetitive examples that explained earlier. In addition to that, four new examples mentioned below and are provided within the context of the text.

Screenshot (4) the Retrieval of 6-grams

Total No. of N-Gram Types			7	Total No. of N-Gram Tokens	
Rank	Freq	Range	N-gram	21	
1	3	1	baltimore state hospital for the criminally		
2	3	1	i d like to talk to		
3	3	1	see if i can find it		
4	3	1	she didn t give a damn		
5	3	1	she didn t look at him		
6	3	1	state hospital for the criminally insane		
7	3	1	the baltimore state hospital for the		

The following new examples are given within the context in addition to explanatory analysis:

After five minutes, she said, "It was strange going in there. Sometimes **I'd like to talk to** you about it."

"Later, **I'd like to talk to** you, Mr. Bimmel. What about Fredrica's mother?"

"Do you know where Mrs. Lippman is? **I'd like to talk to** her."

The above examples are related to Clarice the main character in the novel. The six clusters sentence "She didn't give a damn" occurs 3 times in the novel.

The pronoun 'she' refers once again to Clarice Starling. The occurrences happen to be in chapter (34) to describe her personality when she was young and how she did not pay attention to boys and clothes and commanding.

The other two occurrences are within the same paragraph "**She didn't give a damn** about commanding. She found **she didn't give a damn**, or a shit for that matter, about being Special Agent Starling."

When Starling had a conversation with Agent Jack Crawford "Starling nodded, kept her head down after the last nod. Maybe he would find it easier to talk if **she didn't look at him**. He was steady, but something had hold of him.

Starling wondered for a second if his wife had died. Or maybe spending all day with Catherine's grieving mother had done it." (Ch.18). When Starling had a conversation with Dr. Lecter "**She didn't look at him**. She –didn't want to match stares now. This was not a confrontation." (Ch.25).

Pilcher nodded. "Yes," he said when **she didn't look at him**." (Ch.40).

From the above clusters, it can be noticed that Starling's attitude while making a conversation with others through awkward situations she prefers to avoid eye contact with the interlocutors.

"Would you shine your light on the clipboard? Let's **see if I can find it** here."

"WEYE mobile unit just turned in at the gate, Joney," the man, Harry, said.

"Let's see **if I can find it** here, Officer, here it is." (Ch. 56).

"Did you know Fredrica Bimmel when she worked for Mrs. Lippman?"

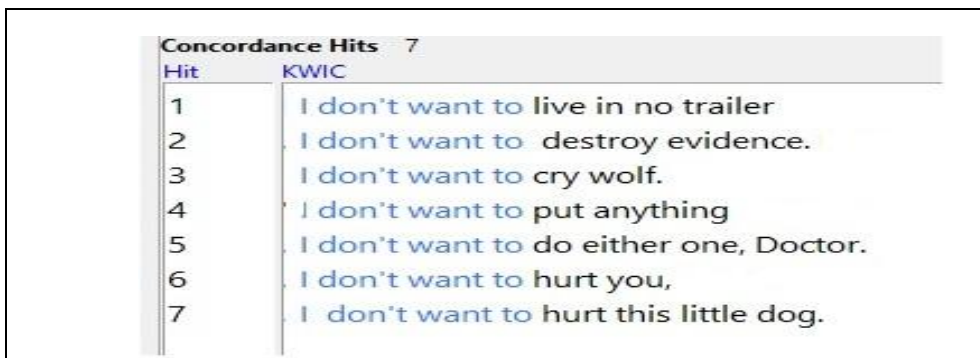
"No. Was she a great, fat person? I may have seen her, I'm not sure. I didn't mean to be rude--- I was sleeping... Mrs. Lippman had a lawyer, I may have his card somewhere, I'll see **if I can find it**" (Ch.53).

5-grams

Malhberg (2013) makes a detailed study of 5-grams and divides five word clusters or lexical bundles into five categories; speech clusters, as if clusters, time and place clusters, body parts clusters and labels (73).

Bujanova (2013) explains that the dominant examples of 5-grams include the pronoun "I". This can give information about the characters' needs or "unwillingness, intentions and inability"(39). For example, " *I don't want to*" in the following image:

Screenshot (5) the Retrieval of 5-grams



The screenshot displays a concordance search interface with the following data:

Concordance Hits 7	
Hit	KWIC
1	I don't want to live in no trailer
2	I don't want to destroy evidence.
3	I don't want to cry wolf.
4	I don't want to put anything
5	I don't want to do either one, Doctor.
6	I don't want to hurt you,
7	I don't want to hurt this little dog.

Conclusions

The study of corpus stylistics can be of great importance for researchers. In other words, it connects both quantitative and qualitative methods, also it helps to bring readers'/ researchers' attention to certain crucial elements that literary texts have.

Moreover, computational tools that are employed to perform the task of analyzing the chosen novel are practical and beneficial not only in terms of saving time and gaining quick results by one click, but also these tools calculate words that might be missed by humans. Yet, these numbered results cannot be significant without a manual study. Thus, the quantitative method plays a completion role to foster the analysis. This is done by shedding light on crucial items, studying them, and connecting their importance in the context of the text. There must be some sort of a sifting technique that separates those guiding words (leading to the theme) from frequent words that have less importance and provide little information about literary texts.

Regarding what has been mentioned, the study reveals many valuable features about the selected novel. It shows how corpus stylistics aids to give prominence to the associations of words as well as to thematic-signal words. For instance, the words "body," "murder," "Starling," and "victim" serve as indicators to the major events of the novel. Other findings from the study show that the occurrences of words help to locate certain information. To be more specific, the frequent

occurrence of the word “body” with the words “water” and “river” for example, illustrates the fact that the criminal gets rid of the victims' bodies by throwing them in the water.

Another interesting fact is that key semantic domains help to put words in groups and label them to the suitable semantic domain. Certainly, without the help of corpus-based study, stylisticians do not pay attention to each and every word, the computer does the job of sorting in an objective, systematic way as well as in a short time.

Finally, studying n-grams plays a significant role in this study. Readers may ask why clusters are important to be examined. These repeated sequences of words are vital to foster the interpretation of any literary text. They lead in a direct way to basic themes and/or ideas. Additionally, they reveal something about the style of the novel. To be more specific, long n-grams have importance over short ones. For example, the 8-gram "Baltimore State Hospital for the Criminally Insane" gives information about the grouping of significant words together, like “hospital” which is associated with basic events and characters of the novel.

To sum up, corpus stylistics proves to be of great help in conducting this study. The chosen methodology works as a perfect guide to understand elements of the text in a logical and systematic way. The resulting findings prove this fact.

References and notes:

- Anthony, Laurence. *AntConc (Version 3.2.4m) Computer Software*. Tokyo, Japan: Waseda University, 2011. Web 18 July 2016, <http://www.antlab.sci.waseda.ac.jp/>
- Baker, Paul, Andrew Hardie and Tony McEnery. *Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2006. Print.
- Bujanova, K. A Corpus-Cylistic Analysis of Mitchell's "Gone with the Wind" and Hemingway's "A Farewell to Arms", 2013. Web 15 June 2016, <https://www.duo.uio.no/bitstream/handle/10852/37048/Bujanova-Masterxx2x.pdf?sequence=2>
- Fialho, Olivia, and Sonia Zyngier. "Quantitative Methodological Approaches to Stylistics." *Routledge Handbooks in English Language Studies*. Ed. M. Burke. Abingdon: Routledge, 2014. 329. Print.
- Gliozzo, Alfio, and Carlo Strapparava. "Semantic domains." *Semantic Domains in Computational Linguistics*. Springer Berlin Heidelberg, 2009. 13-32.
- Harris, Thomas. *The Silence of the Lambs*. Web 10 May 2016, <http://www.gyxsqex.com/tushu/book/book56/20098306242216.pdf>
- Hoey, Michael. *Text, Discourse and Corpora: Theory and Analysis*. A&C Black, 2007. Print.

- Huntley, Chris. *Dramatica. The Silence of the Lambs*. n.d. Web 14 June 2016, <http://dramatica.com/analysis/the-silence-of-the-lambs#os>
- Jabbar, Nadia. *A Corpus Stylistic Analysis of Selected Plays of the Theatre of Anger by John Osborne*. MA thesis. University of Baghdad, Baghdad, Iraq, 2014.
- Jaafar, Eman Adil. "A Stylistic Analysis of Bisson's Bears Discover Fire." *Journal of College of Education for Women* ISSN 1680-8730, 25.4.2014: 1173-1182.
- Jaafar, Eman Adil. "Examining the Language of Drama texts with a Reference to Two Plays: A stylistic study". *International Journal of Humanities and Cultural Studies (IJHCS)* ISSN 2356-5926, 3.1 .2016:564-575.
- Jeffries, Lesley, and Daniel McIntyre. *Stylistics*. Cambridge: Cambridge University Press, 2010. Print.
- Leech, Geoffrey N., and Mick Short. *Style in Fiction*. London & New York: Longman,1981.Print.
- Mahlberg, Michaela. "Corpus Stylistics". *The Routledge Handbook of Stylistics*. Ed M. Burke. Abingdon: Routledge, 2014. 378-392. Print.
- Mahlberg, Michaela. *Corpus Stylistics and Dickens's Fiction*. London: Routledge, 2013. Print.
- Mahlberg, Michaela. & McIntyre, Dan. A Case for Corpus Stylistics: Ian Fleming's Casino Royale. *English Text Construction*, 4.2 .2011: 204-227.
- McIntyre, Dan. "Towards an Integrated Corpus Stylistics." *Topics in Linguistics* 16.1 .2015: 59-68.
- Rayson, Paul. *Matrix: A Statistical Method and a Software Tool for Linguistic Analysis through Corpus Comparison* .Diss. Lancaster University,2003.Web 4 June 2016.
- Rayson, Paul. 'From keywords to key semantic domains', *International Journal of Corpus Linguistics*, 13 (4): 519–549.2008. Web 11 July 2016, <http://ucrel.lancs.ac.uk/wmatrix>
- Rayson, Paul. "WmatrixCorpus Analysis and Comparison Tool". Web 16 June 2016 <http://ucrel.lancs.ac.uk/wmatrix/>
- Semino, Elena, and Mick Short. *Corpus stylistics: Speech, Writing and Thought presentation in a corpus of English writing*. London: Routledge, 2004. Print.
- Trier, Jost. "Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes, I: Von den Anfängen bis zum Beginn des 13." *Jahrhunderts*. Heidelberg ,1931. Print.
- Wales, Katie. *A Dictionary of Stylistics*. London: Routledge, 2014. Print.

Summary

Corpus Stylistic Analysis of Thomas Harris' *The Silence of the Lambs*

Eman Adil Jaafar

University of Baghdad, Baghdad, Iraq

This paper aims to employ one of the corpus stylistic methods to analyze Thomas Harris's novel, *The Silence of the Lambs*. Recently, technology has invaded our lives. To put it differently, researchers depend highly on computers to access and gain information about certain data. Thus, it is crucial to keep up with the up-to-date developments concerning computational methodologies and toolkits. Corpus stylistics helps to find certain features that cannot be understood without using the techniques of computers. In order to achieve this goal, a quantitative and qualitative methodology is applied. Corpus stylistics helps to analyze lengthy texts more efficiently. This is not to say that it substitutes the manual stylistic one. In fact, both the corpus and manual stylistic analyses work hand in glove, and they complement each other. The tool that is used to conduct the analysis by examining keywords and key semantic domains is Wmatrix3. In addition to the previous tool, AntConc is a complementary tool to investigate n-grams in the novel and to point out their significance to the overall interpretation.

Keywords: Corpus stylistics, Wmatrix3, AntConc, N-grams, Quantitative & Qualitative approach, Keywords, Semantic domains