Pakistan Languages and Humanities Review
www.plhr.org.pk

RESEARCH PAPER

Stylometric Appraisal of Coelho's Selected Translated Novels: A Digital Humanities Study

Dr. Zafar Ullah*1 Dr.SundasTahreem2 Samira Amjad3

1. Assistant Professor, English Department, University of Education, Vehari, Punjab, Pakistan
2. Assistant Professor, Muslim Youth University, Islamabad
3. English Teacher, Govt. Girls Primary School, Jhatta, Sialkot, Punjab, Pakistan

| *Corresponding Author | zafarullah@ue.edu.pk |

ABSTRACT

To quantify the style markers of any writer is a laborious and erroneous task in close reading, while distant reading with the Voyant Summary tool presents stylometric features precisely. The current study employs the Voyant summary tool to analyse data through the lens of Stefan Sinclair and Geoffery Rockwell's Hermeneutica Theory (2016). The Voyant Summary tool performs text mining of Coelho's style markers; total words, unique words, vocabulary density, the average length of sentences and the most frequent words. Coelho's vocabulary density ranges from 0.075 to 0.263 words. Colhelo's average length of sentences ranges from 13 to 26.2 words per sentence. The dialogic writing style is evident with the word 'said'. Moreover, themes of love and life are the most repeated ones. To enhance the readability of his works, he refrains from archaic vocabulary. The Zahir shows the lowest vocabulary density, while Adultery reveals the highest vocabulary density.

Introduction

The main focus of stylometry is to explore distinct stylistic features which define a writer's particular writing style (Rybicki, Hoover, & Eder, 2016). Stylometry or computational stylistics utilises computers to analyse texts from different perspectives (Rybicki, Hoover, & Eder, 2016) and studies an author's style with a quantitative approach (Eder, Rybicki, &Kestemont, 2016). Stylistics focuses on the 'style markers', such as the occurrence of the words, the appearance of punctuation marks, the usual length of the sentences, and the regular usage of words (Eder, Piasecki, &Walkowiak, 2017). In modern studies, documents are studied with different digital tools which find the 'stylistic fingerprint' of an author (Eder, Piasecki, &Walkowiak, 2017). The Voyant Summary Tool was designed by Sinclair & Rockwell (2016) in Canada. This digital tool is capable of highlighting the style markers of different authors (Sinclair & Rockwell, 2015). The current study finds the literary stylistic fingerprints of Paulo Coelho's 15 selected novels through the Voyant Summary tool. These novels are *The Pilgrimage, The Alchemist, The Valkyries, Aleph, By the River Piedra: I Sat Down and Wept, The Fifth Mountain, Veronika Decides to Die, The Devil and Miss Prym, Eleven Minutes, The Zahir, The Witch of Portobello, The Winner Stands Alone, The Supreme Gift,* and *Adultery*. All these novels are translated from Portuguese to English. In the current study, the Voyant Summary tool reveals total words, unique words, vocabulary density, the average

length of sentences, and the Most Frequents Words (M.F.Ws). This study delimits the ten most frequent words from each novel.

Literature Review

Stylometry had deep roots in the past. Due to the extraction of textual features from large texts, it was vital to use computerised tools to examine and quantify text, making it better for the readers' understanding (Rybicki, Hoover, & Eder, 2016). Digital humanities defined the style on a more extensive level than before (Hermann, van Dalen-Oskan&Schoch, as cited in O'Sullivan et al., 2018). Stylometry filled the gap between computer science and humanities. Different methods of stylometry were being used in the Humanities (Eder, Rybicki, &Kestemont, 2016). In humanities, computational stylistics found the differences and similarities in stylistic patterns of different writers (Eder, Rybicki, &Kestemont, 2016). Advanced information could be gained through the use of technology, which saved researchers' time and produced authentic results even if the text had yet to be read minutely (Ullah, Uzair, & Mahmood, 2019).

The style was the most important factor of stylistics; it was the pattern of a writer composing his work (Rybicki, Hoover, & Eder, 2016). The style came under the category of linguistic features, but sometimes other non-linguistic features in a text contribute to a style. There were patterns and distributions, but reoccurrence and regularity were the most important elements of a style (Rybicki, Hoover, & Eder, 2016). The reoccurrence of function lexemes such as any, of, a and an in documents determined the writer's style (Amancio, 2015). Computational stylistics favoured qualitative research, but it was important in computational stylistics to find the quantitative elements associated with the style (Eder, Rybicki, &Kestmont, 2016).

Stylometry's major role was to analyse a text from different perspectives. In the beginning, it was used to identify the author of any piece of writing based on the style (Eder, Rybicki, &Kestmont, 2016). Mostly, different ways were used to distinguish one author from others (Rybicki, Hoover, & Eder, 2016). Computational stylistics was first known for authorship attribution, but later it was expanded to different analyses which recovered different linguistic elements and literary styles (Eder, Piasecki, &Walkowiak, 2017). Computational stylistics was concerned with the style of a writer in documents and with their meta-data (Eder, Rybicki, &Kestemont, 2016). Authorship studies were the most popular application of stylometry from the literary point of view (Eder, Piasecki, &Walkowiak, 2017). Due to stylometry, a large number of readers from all around the world were attracted towards authorship attribution, as it had been using authentic computational soft wares to find it (Eder, Rybicki, &Kestemont, 2016). Stylometry attempted to explore large text collections using computational techniques (Eder, Piasecki, &Walkowiak, 2017). Although stylometry was not for replacing one author with others, yet it was used to make reading beneficial and understandable for readers and scholars (Rybicki, Hoover, & Eder, 2018). The purpose of stylometry was not to replace previous knowledge with new information; rather, new information could easily be fitted into the developing process with the help of computational stylistics (Eder, Rybicki, &Kestemont, 2016). Stylometry chiefly involved comparison; one's style was compared to the others. The manual and digital electronic analysis eased comparative stylistic analysis (Rybicki, Hoover, & Eder, 2016).

The computational approach was mostly used to analyse large text or corpus. Computational stylistics could be conducted on a definite set of analyses; more work could be done if stylometry techniques were accessible to researchers (Eder, Piasecki,

&Walkowiak, 2017). 'High-level graphical user interface' analyses were done with the help of stylometry with R, as new researchers and scholars could equally get help from it. Digital Humanities tools and software contributed to facilitating researchers for efficient and accurate computation (Eder, Rybicki, &Kestmont, 2016). The computational stylistic system depended on 'server installation combined with a web-based user interface.' This combination eliminated the chances of installing any extra application. Several applications that needed to combine the texts to distinguish among their feature were conducted through the stylistic process. The major hurdle in using computational stylistic procedures was deficient information and competence because it required additional software whose usage was complicated for humanities researchers (Eder, Piasecki, &Walkowiak, 2017). The use of stylometry might be challenging as the findings were based on counted lexemes from documents. In contrast, the distribution was not according to the counted lexemes but rather on the authorship attribute (Waugh, Adams, & Tweedie, 2000).

Stylometry was more helpful for the problems at three levels; 'microanalysis, middle-distance analysis, and macro analysis or distant reading (Rybicki, Hoover, & Eder, 2016). In stylometry, token words were as important as the other meaningful lexical items while analysing. The most frequent words were analysed to find the sameness between collections of texts (O'Sullivan, 2018). Stylometry was more concerned with words as the order of words and their occurrence conveyed a particular theme and effect (Rybicki, Hoover, & Eder, 2016). Specific characteristics in a text were statistically analysed by the traditional patterns (Amancio, 2015). Most of the textual features could only be analysed through computational approaches. It was a fact that in any document, words were the most frequent elements and also easier to identify and count as compared to other textual characteristics (Rybicki, Hoover, & Eder, 2016). Complicated documents were analysed through stylometry; all the languages were not equally suitable due to the different linguistic structures. Identification of the most frequent words was easy in English than in the inflectional languages (Eder, Piasecki, &Walkowiak, 2017). One study was conducted on textual and stylistics analysis of 1000 novels from Jonathan Swift to E.L. James to find the lexical frequencies; an evident difference was observed (Rybicki, Hoover, & Eder, 2016). The stylometry of short stories was studied with the help of the Summary tool in the Voyant suite. This study explored total words, unique words, vocabulary density and the average length of sentences (Ullah, & Mahmood, 2019).

Several research works were conducted manually on Paulo Coelho and his writing style. Leech and Short's (2007)stylistic analysis theory applied to the Alchemist to reveal lexical categories and grammatical categories which manifest his style. The novelist used techniques of foreshadowing, irony, personification, symbolism, motifs, imagery, metaphors, and symbols with the simplicity of words. 47% concrete nouns, 53% abstract nouns and 23% proper nouns have been used. Moreover, 43% of stative verbs, 34% of dynamic verbs, 49% of transitive verbs, 29% of intransitive verbs and 27% of copula verbs have been used (Sarfraz, 2022). Besides, Coelho's personal life had a great impact on his writing. Most of his works were autobiographical. His second novel, *The Alchemist* (1988),was an achievement for him as an international writer (Mayer & Maree, 2017). Magic laid a significant impact on Paulo's works (Mayer & Maree, 2018). Most of Paulo's works had life-changing effects on readers' lives (Manogna, Charyulu, & Kumar, 2018).

This study selected 100 novels written between 1920 to 1940 for statistical analysis of the literary style. (Modoc, &Gârdan, 2020). Wuxia, a translation of Chinese fiction into English and Eastern, as well as Western hero-related literature, manifested

some shared stylistic features. To conduct the study, a stylistic panorama has been applied. The study revealed that the translation of Wuxia was mismatched with the Western subgenre in terms of most frequent words and sequence. Wuxia is different in vocabulary and the terseness of paragraphs (Wu, & Li, 2022).

Several research works have been conducted using the Voyant tool. In this article, the Voyant Summary tool has been used. There was a large quantity of existing knowledge about computational stylistic analysis of different authors from all around the world that helped the new scholars and researchers to get help and ideas for new research works in the stylometry domain. This study stylistically analysed the 15 selected novels of Coelho using the Voyant Summary tool, which would be helpful for new learners to get an idea about the stylistic fingerprint of Paulo Coelho and would contribute to further studies on any writer in this respected domain.

Material and Methods

The current research implemented the following parameters to accomplish the research process.

Research Model

This research followed the descriptive method; it was based on both qualitative and quantitative analysis of the selected novels. The tabular data comprised words and numbers; hence, they were analysed with a mixed methods approach.

Research Design

Stefan Sinclair and Geoffery Rockwell's Hermeneutica Theory (2016) was applied to analyse Paulo Coelho's 15 selected novels.

Research Tool

The Voyant Summary is one of the 19 tools from the text mining suite, which was designed by Stefan Sinclair and Geoffery Rockwell.

Data Required

The current study built a corpus of Paulo Coelho's fifteen selected novels.

Data Collection

The data were collected through the Voyant Summary tool, which had style markers such as total words, unique words, vocabulary density, the average length of the sentences and the ten most frequent words of each novel.

Data Analysis

Data were analysed with the qualitative approach with the Hermenentica theory propounded by Sinclair and Rockwell (2016).

Theoretical Framework

Stefan Sinclair and Geoffrey Rockwell (2016) propounded Hermeneutica Theory while interpreting Voyant tools. Key postulates of the theory were:

i. Hermeneutica Theory is embedded in a context.
ii. It is not like black boxes.
iii. Manipulation is in service of exploration and understanding.
iv. It is supplemented by other materials.
v. Knowledge-bearing tools provoke reflection.
vi. Hermeneutic tools fail in interesting ways.
vii. They can be extended to expose new things (Rockwell & Sinclair, 2016, p. 166).

Results and Discussion

Stylometric appraisal of Coelho's fifteen selected novels has been conducted with the help of text mining tools that inform the readers about total words, unique words, vocabulary density, the total length of sentences and the ten most frequent words. Quantitative and qualitative data have been attached to the analysis of each novel separately.

The Pilgrimage (1987)

This corpus has 1 document with 66,711 total words and 5,746 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.086

Average Words Per Sentence: 17.0

Most **frequent words** in the corpus: petrus (390); said (293); road (239); time (184); began (172); sword (152); going (134); santiago (123); love (110); way (104)
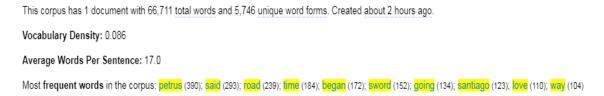
Figure 1: *The Pilgrimage* through Voyant Summary tool

This novel contains 66,711 total words, with 5,746 unique words which are used more than 11 times throughout the novel. The vocabulary density of the novel is 0.086, which is obtained by dividing unique words by total words. The summary tool also reports the average length of the sentences, which is 17.0 in this novel which shows that the writer has written medium-length of sentences. The reoccurrence of words and sentence length makes the novel understandable for the readers. The most frequent words of any novel reveal key themes to inform about the plot of the novel. Petrus (390) and Santiago (123) are the proper nouns, and the former guides the boy. Other most frequent words are road (239), going (134), time (184), and way (104). These words represented the time of the novel when most of the travelling was done through roads. The plot of novel revolves around a boy who travels from North Spain to Santiago. Sward (152) represents his award which he gets in the end. The main theme of the novel is love (110). The word said (293) indicates the writer's dialogic and direct style.

The Alchemist (1988)

This corpus has 1 document with 38,900 total words and 3,323 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.085

Average Words Per Sentence: 14.4

Most **frequent words** in the corpus: boy (498); said (315); desert (179); man (164); alchemist (148); old (129); know (112); world (111); thought (101); sheep (97)
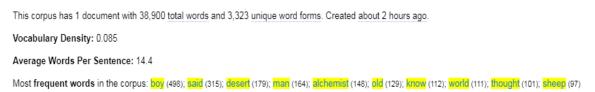
Figure 2: *The Alchemist* through Voyant Summary tool

There are 3,323 unique words in this novel which have reappeared more than 11 times, so the total words are 38,900 with a vocabulary density of 0.085, which is almost

the same as the first novel. The average length of the sentences is 14.4, which makes it easier for intermediate-level readers to comprehend the text as it is short. The most frequent words show the stylistic features of a writer. The most frequent words were analysed to find the sameness between collections of texts (O'Sullivan, 2018). The most frequent words demonstrate the whole plot and the main theme of the novel. The main character, boy (498), is a shepherd at the beginning and has sheep (97), but he wants to travel all around the world (111), and later, he travels through the desert (179) to follow his dream, so there he met with a man (111) named Alchemist (148). The words like thought (101) and know (112) reveal the boy's character that he has an ambitious and curious nature. The word old (129) informs the readers that the novel has some historical instances in it. Again the word 'said' has been used 315 times which reveals Coelho's dialogic writing style.

Brida (1990)

This corpus has 1 document with 58,063 total words and 5,065 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.087

Average Words Per Sentence: 16.4

Most **frequent words** in the corpus: brida (367); said (296); wicca (237); world (162); magus (160); know (159); time (156); love (149); tradition (148); people (146)
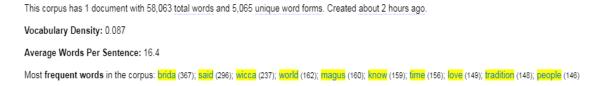
Figure 3: *Brida* through Voyant Summary tool

The novel '*Brida'* consists of 58,063 total words while having 5,065 unique words, and 0.087 is its vocabulary density. Unique words are used more than ten times by the writer throughout the novel. Length sentences are 16.4 per sentence. In '*Brida'*, most of the frequent words consist of proper and common nouns; Brida (367), Wicca (237), and Magus (160) are the names of the main characters. World (162), time (156), love (149), traditions (148), and people (146) are the common nouns. The major theme is love, and the direct pattern of writing is revealed through the use of said (296). This novel is about a girl named Brida who is too curious about knowledge and deeply interested in magic. She finds a man and woman who teach her magic. Gradually, she learns how to maintain a balance between her belonging and her wish to become a witch. This novel concentrates on motifs of love, passion, and spirituality.

The Supreme Gift (1991)

This corpus has 1 document with 8,568 total words and 2,060 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.240

Average Words Per Sentence: 26.2

Most **frequent words** in the corpus: love (182); life (36); things (34); world (30); god (26); man (25); people (21); know (20); great (19); faith (18)
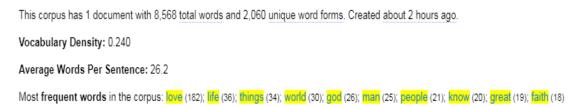
Figure 4: *The Supreme Gift* through Voyant Summary tool

This novel has 8,568 total words, having 2,060 unique words which have been used four times in the novel, so its vocabulary density is 0.24. The writer has used quite long sentences, as the average length of sentences is 26.2. This novel has fewer words but longer sentences as compared to the first two novels. The most frequent words are the key themes of any writing. In this novel, the words like the world (30), man (25), know (20), and people (21) represent real life, where the characters and the situation is real and natural. The other words, god (26), great (19), and faith (18) communicate a

spiritual sensation in the novel. Life (36) and love (182) represent the key themes of the novel.

The Valkyries (1992)

This corpus has 1 document with 42,666 total words and 6,364 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.149

Average Words Per Sentence: 13.0

Most **frequent words** in the corpus: paulo (381); said (267); chris (194); valkyries (158); angel (153); desert (122); know (118); coelho (112); valhalla (111); angels (100)

Figure 5: *The Valkyries* through Voyant Summary tool

In *the Valkyries*, the total number of words is 42,666, and the unique words are 6,364, which are repeated seven times in the whole novel, 0.149 is the vocabulary density of the Valkyries. The most frequent words of this novel reveal the subject matter and also the main characters. As in this novel, proper noun as Paulo (381), Coelho (112), Chris (194), Valkyries (158), and Valhalla (111) are used which inform readers about the main characters. Paulo Coelho is a major character, Chris is his wife, and Valkyries is the subject matter of the novel, which is a group of warrior women. Valhalla is the leader of that group of Valkyries. Desert (112) shows the setting of the novel, and the words like angle (153) and angles (100) represent the presence of supernatural elements in the novel. As Paulo himself has said, "I wrote a book about angels; I believe in angles. This book is called the Valkyries." Again, the word know (118) informs about the characters' nature and that they are well aware of their surroundings. Like other novels, the word said (267) demonstrates the dialogic writing style of the writer. The average length of sentences in the Valkyries is 13.0, which indicates that Coelho's writing style is simple because long sentences make reading and comprehension a difficult task, whereas Coelho uses medium-length sentences.

By the River Piedra: I Sat Down and Wept (1994)

This corpus has 1 document with 40,252 total words and 3,735 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.093

Average Words Per Sentence: 12.3

Most **frequent words** in the corpus: said (310); love (188); know (119); life (106); woman (102); time (99); god (95); want (87); man (86); asked (83)

Figure 6: *By the River Piedra: I Sat Down and Wept* through Voyant Summary tool

The writer has used 3,735 unique words almost eleven times, as the total words are 40,252. So the vocabulary density is 0.093. As every writer uses his/her specific stylistic patterns. Coelho uses easy diction; most of his frequent words are nouns. As in this novel, the most frequent words are the woman (102) and man (86), revealing the main characters of the novel. Then the other words are love (188), life (106), time (99), want (87), and know (119), which inform about the theme as well as the real-life setting of the novel. The whole plot of the novel revolves around the life of a woman who has been struggling to find out the real meaning of his life. Then she meets a man who is her childhood friend after a long time. He helps her to find his true self and a strong connection with god, and they both find their true love in one another. Again, words said (310) and asked (83) suggest the dialogic nature of Coelho's writing style. The

average length of sentences in this novel is 12.3, which makes it easy to read and understand. The novel revolves around a girl named Pilar. She faces many problems in her life and then reunites with one of his childhood friends. Eventually, she realises that he is the love of her life. Later, she comes to know that her love has to choose one thing; either her or his spiritual callings, but in the end, due to their true love and struggles, they manage to be united. This novel is about love and the miracles of love.

The Fifth Mountain (1996)

This corpus has 1 document with 53,191 total words and 4,651 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.087

Average Words Per Sentence: 14.4

Most **frequent words** in the corpus: elijah (420); said (263); lord (180); akbar (174); city (156); governor (142); time (136); boy (135); god (135); people (133)
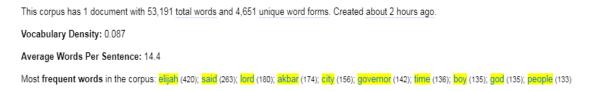
Figure 7: *The Fifth Mountain* through Voyant Summary Tool

In this novel, 4,651 unique words have been used that reappeared eleven times with a vocabulary density of 0,087 which is almost the same as in the Alchemist and the Pilgrimage. The average length of sentences used by the writer is 14.4. Similar to the other novels, the most frequent words are nouns; Elijah (420), lord (180), Akbar (174), city (156), governor (142), boy (135), god (135), and people (133).

The plot of this novel revolves around a man whose name is Elijah, he is the prophet of god and lives in Israel, but because of some reason, he has to move to a town named Akbar. The plot proceeds and time (136) comes, he becomes the governor of the town. In the novel, Elijah is guided by the angels of the lord. Like the other novels, dialogic patterns are used by Coelho as said (263) is the second most frequent word.

Veronika Decides to Die (1998)

This corpus has 1 document with 53,247 total words and 7,010 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.132

Average Words Per Sentence: 17.3

Most **frequent words** in the corpus: said (209); veronika (195); people (162); time (148); life (146); eduard (130); just (126); know (107); mari (106); dr (105)
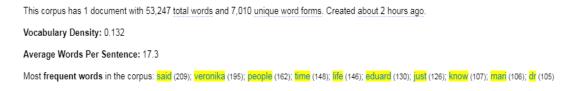
Figure 8: *Veronika Decides to Die* through Voyant Summary Tool

This novel comprises 53,247 total words, among which 7,010 words are unique, and the vocabulary density is 0.132. The writer uses unique words more than seven times in the whole novel. The writer has used nouns as the most frequent words, such as Veronica (195), people (162), Eduard (130), Mari (106), and dr (105), revealing the main characters of the novel. Time (148) and life (146) elucidate the major themes of the novel. Most frequent words like just (126) and know (107) inform the readers about the nature of the novel as, throughout the novel, characters are trying to find out the meaning and purpose of their lives. The use of direct narration can be seen in the word said (209). As the most frequent words suggest that Veronica is the main character of the novel and that the plot of the whole novel revolves around her. In the beginning, she tries to end her life but somehow goes to a mental hospital, where she gets treatment for her psychological disorders. There are other patients like Eduard and Mari. So, at the end of the novel, Veronica finds the meaning of her life and true love as well in Eduard. The major themes of the novel are time and life. With the passage of time in the

hospital, she finds the solution to her problem and gets the time to think about them, which changes the meaning of life for her.

The Devil and Miss Prym (2000)

This corpus has 1 document with 47,957 total words and 5,024 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.105

Average Words Per Sentence: 20.4

Most **frequent words** in the corpus: said (182); village (148); gold (135); viscos (135); stranger (133); know (131); time (127); chantal (124); man (120); people (118)
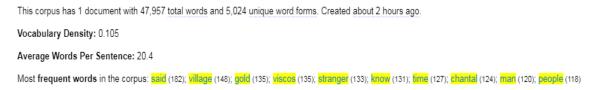
Figure 9: *The Devil and Miss Prym* through Voyant Summary tool

The writer has 47,957 total words, among which 5,024 are unique words. These unique words have been used almost nine times. The vocabulary density of this novel is 0.105. In this novel, the writer has written long sentences. The average length of sentences is 20.4. Like other novels, this novel also has nouns as the most frequent words; Viscos (135), Chantal (124), stranger (133), man (120), people (118), and village (148). The main characters and setting of the novel reveal through the most frequent words. Gold (135) and time (127) suggest that materialism is the main theme of the novel. Stylistic markers exhibit the main themes (Ullah, Uzai, & Mahmood, 2019). Again know (131) and said (182) demonstrate the character's curious nature and the dialogic pattern of writing.

The setting of the novel is a village named Viscos, where lives a woman who is offered money for killing someone. Throughout the novel, people are running behind worldly desire, but in the end, the woman realises that it is wrong and that being greedy can be dangerous for them.
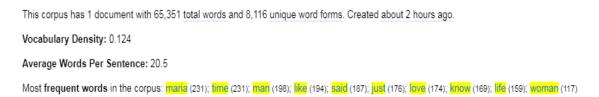
Eleven Minutes (2003)

This corpus has 1 document with 65,351 total words and 8,116 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.124

Average Words Per Sentence: 20.5

Most **frequent words** in the corpus: maria (231); time (231); man (198); like (194); said (187); just (176); love (174); know (169); life (159); woman (117)

Figure 10: *Eleven Minutes* through Voyant Summary tool

'*Eleven Minutes*' contains 8,116 unique words; the writer has used these words eight times to write this novel; total words are 65,351 with 0.124 vocabulary density, 20.5 is the average length of sentences. So the writer has composed quite long sentences, as in *the Devil and the Miss Prym*. The most frequent words in this novel are Maria (231), time (231), man (198), and woman (117). The use of the word like (194) shows that the writer has used many similes to refer to different things. Then the major theme is love (174) and life (159). Direct narration can be seen through the use of said (187). The other most frequent words are just (176) and know (169). *Eleven Minutes* revolves around a lady named Maria. At the beginning of the novel, she tries to get fame and money by doing wrong deeds. On the other hand, throughout the novel, she struggles to find out the essence of true love, and at the end of the novel, she attempts to discover her inner light.

The Zahir (2005)

This corpus has 1 document with 86,124 total words and 6,483 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.075

Average Words Per Sentence: 17.9

Most **frequent words** in the corpus: know (308); people (289); love (279); i'm (240); just (230); time (229); don't (215); like (200); said (189); life (187)

Figure 11: *The Zahir* through Voyant Summary tool

'*The Zahir*' has 86,124 total words having 6,483 unique words, and the writer has used these words more than thirteen times to write this novel. Medium-length sentences have been used in this novel as the average length of sentences is 17.9, which is easy to be comprehend by readers. Most frequent words reveal the writer's style, main themes and characters. This novel has been written in the first person singular narration as the use of I'm (240) and don't (215) suggest it. Like (200) again informs about the use of similes. Again, the prominent themes are love (279), time (229) and life (187). Know (308), just (230), and people (289) suggest that the novel is about real life. Said (189) indicates the use of a direct pattern. The novel is about a wealthy young man whose wife escapes with a friend, and he tries to find her out. As the plot proceeds, he finds the meaning of true love. This novel is about following your dreams or destroying them.

The Witch of Portobello (2006)

This corpus has 1 document with 79,841 total words and 6,943 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.087

Average Words Per Sentence: 18.0

Most **frequent words** in the corpus: athena (248); said (221); know (207); people (206); love (183); don't (179); i'm (179); time (168); like (158); life (144)

Figure 12: *The Witch of Portobello* through Voyant Summary tool

'*The Witch of Portobello's*' vocabulary density is 0.087 as the total words are 79,841 and the unique words are 6,943. The unique words are used more than eleven times by the writer. The average length of sentences in this novel is 18, which is neither too long nor too short. Same as the Zahir, the Witch of Portobello is also in the first person singular narration, as I'm (179) and don't (179) are the most frequent words in the novel. The writer has used both proper and common nouns such as Athena (248) and people (206). Like other novels, the word "like (158)" is also used in this novel. Major themes which are suggested by the most frequent words are love (183), time (167), and life (144). Dialogic narration is shown by said (221), and the characters' desires to know more are shown by know (207).

The Winner Stands Alone (2008)

This corpus has 1 document with 122,785 total words and 9,635 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.078

Average Words Per Sentence: 19.2

Most **frequent words** in the corpus: people (351); just (302); know (281); world (261); like (259); time (249); man (241); he's (213); life (199); it's (194)

Figure 13: *The Winner StandsAlone* through Voyant Summary tool

Among these fifteen novels, '*The Winner Stands Alone*' contains total words which are 1,22,785, and the unique words in this novel are 9,635; as a result, vocabulary density is 0.078, which means that the writer has used these unique words for more than twelve times. In the case of the average length of sentences, the writer has used long sentences, as the average length is 19.2 in this novel. If we talk about the most frequent words, so in this novel, the most frequent words show that the writer has used both third person narration in this novel as words like he's (213), and it's (194) are used. People (351), man (241), life (199), the world (261), and time (249) are the nouns which reveal the characters and major themes of the novel. Other most frequent words are; know (281), just (302) and like (259). It is a tale of a rich Russian millionaire. A struggle for fame and fortune is there throughout the novel.
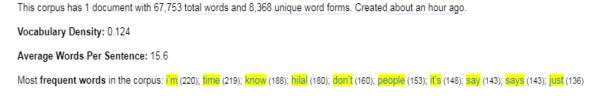
Aleph (2010)

This corpus has 1 document with 67,753 total words and 8,368 unique word forms. Created about an hour ago.

Vocabulary Density: 0.124

Average Words Per Sentence: 15.6

Most **frequent words** in the corpus: i'm (220); time (219); know (188); hilal (180); don't (160); people (153); it's (148); say (143); says (143); just (136)

Figure 14: *Aleph* through Voyant Summary tool

'*Aleph*' by Coelho contains 67,753 total words while having 8,368 unique words in it that have been used eight times, so the vocabulary density of this novel is 0.124. Reoccurring words make the text easy to comprehend. The writer has used a medium length of sentences in this novel as the average words per sentence are 15.6. Most frequent words in this novel reveal that this novel is in first person and third person narration. The words like I'm (220), don't (160), and the words "it's (148)" inform the narration. Hilal (180) is one of the main characters. The main theme of this novel is time (219). The dialogic style of Paulo Coelho can also be seen in this novel as well. He uses words like say (143) and says (143). The word know (188) demonstrates the curious nature of the characters. Other most frequent words are people (153) and just (136). All these styles markers of Coelho in it suggest that this novel is not much difficult to understand for the readers.

This novel is the autobiography of Paulo himself. In this novel, he talks about his experiences of travelling to different continents. Hilal is the female character who joins him during his journeys. During his journey, he comes to know about his past.
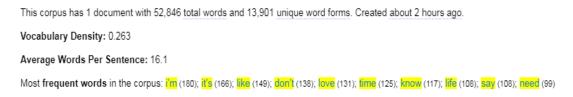
Adultery (2014)

This corpus has 1 document with 52,846 total words and 13,901 unique word forms. Created about 2 hours ago.

Vocabulary Density: 0.263

Average Words Per Sentence: 16.1

Most **frequent words** in the corpus: i'm (180); it's (166); like (149); don't (138); love (131); time (125); know (117); life (108); say (108); need (99)

Figure 15: *Adultery* through Voyant Summary tool

'*Adultery*' novel is comprised of 52,846 total words, among which 13,901 are unique words, and the vocabulary density of this novel is 0.263. Unique words have reappeared almost four times. Like the novel '*Brida*' (1990), this novel also has medium-length sentences of 16.1. In *Adultery* (2014), the writer has used both first and third-

person narration; I'm (180), don't (138), and it's (166) demonstrate the use of both narration styles. Love (131), time (125), and life (149) are the major themes of this novel. Other most frequent words are like (149), know (117), say (108), and need (99). This novel is about a thirty years old lady who is happily married to a wealthy man. She has pride in her fortune, but eventually, she tries to feel that something is missing in her life. Then, she tries to find out that missing part and discover herself. A quantitative analysis of 15 selected novels by Paulo Coelho has been presented in tabular form to give a clear overview of the data.

Table 1
Qualitative data of selected novels collected through the Voyant summary tool

No	Novels	Total Words	Unique Words	Vocabulary Density
01	*The Pilgrimage*	66,711	5,746	0.086
02	*The Alchemist*	38,900	3,323	0.085
03	*Brida*	58,063	5,065	0.087
04	*The Supreme Gift*	8,568	2,060	0.24
05	*The Valkyries*	42,666	6,364	0.149
06	*By the River Piedra: I Sat Down and Wept*	40,252	3,735	0.093
07	*The Fifth Mountain*	53,191	4,651	0.087
08	*Veronica Decides to Die*	53,247	7,010	0.132
09	*The Devil and Miss Prym*	47,957	5,024	0.105
10	*Eleven Minutes*	65,351	8,116	0.124
11	*The Zahir*	86,124	6,483	0.075
12	*The Witch of Portobello*	79,841	6,943	0.087
13	*The Winner Stands Alone*	122,785	9,635	0.078
14	*Aleph*	67,753	8,368	0.124
15	*Adultery*	52,846	13,901	0.263

In table 1, total words, unique words and vocabulary density have been mentioned in the form of qualitative data. From the aforementioned data, it is clear that the vocabulary density of these selected novels is from 0.075 to 0.263. The Zahir is with the lowest vocabulary density, and Adultery is with the highest vocabulary density.

Table 2
Quantitative data of 15 selected novels through the Voyant Summary tool

NO.	Novels	Length of Sentences
01	*The Pilgrimage*	17
02	*The Alchemist*	14.4
03	*Brida*	16.4
04	*The Supreme Gift*	26.2
05	*The Valkyries*	13
06	*By the River Piedra: I Sat Down and Wept*	12.3
07	*The Fifth Mountain*	14.4
08	*Veronika Decides to Die*	17.3
09	*The Devil and Miss Prym*	20.4
10	*Eleven Minutes*	20.5
11	*The Zahir*	17.9
12	*The Witch of Portobello*	18
13	*The Winner Stands Alone*	19.2
14	*Aleph*	15.6
15	*Adultery*	16.1

Table 2 shows the average length of words per sentence. According to the data that are collected through the Voyant Summary tool, the average length of sentences is from 12.3 to 26.2. Mostly, Coelho uses medium sentences as results show that novels often lie from 14 to 20 words per sentence.

Table 3
Qualitative data of 15 selected novels through the Voyant Summary tool

NO.	Novels	The Most Frequent Words
01	*The Pilgrimage*	Petrus (390); said (293); road (239); time (184); began (172); sword (152); going (134); Santiago (123); love (110); way (104)
02	*The Alchemist*	Boy (498); said (315); desert (179); man (164); Alchemist (148); old (129); know (112); world (111); thought (101); sheep (97).
03	*Brida*	Brida (367); said (296); wicca (237); world (162); Magus (160); know (159); time (156); love (149); tradition (148); people (146).
04	*The Supreme Gift*	Love (182); life (36); things (34); world (30); god (26); man (25); people (21); know (20); great (19); faith (18).
05	*The Valkyries*	Paulo (381); said (267); Chris (194); Valkyries (158); angel (153): desert (122); know (118); Coelho (112); Valhalla (111); angels (100).
06	*By the River Piedra: I Sat Down and Wept*	Said (310); love (188); know (119); life (106); woman (102); time (99); god (95); want (87); man (86); asked (83).
07	*The Fifth Mountain*	Elijah (420); said (263); lord (180); Akbar (174); city (156); governor (142); time (136); boy (135); god (135); people (133).
08	*Veronika Decides to Die*	Said (209); veronica (195); people (162); time (148); life (146); Eduard (130); just (126); know (107); Mari (106); dr (105).
09	*The Devil and Miss Prym*	Said (182); village (148); gold (135); viscos (135); stranger (133); know (131); time (127); Chantal (124); man (120); people (118).
10	*Eleven Minutes*	Maria (231); time (231); man (198); like (194); said (187); just (176); love (174); know (169); life (159); woman (117).
11	*The Zahir*	Know (308); people (289); love (279); I'm (240); just (230); time (229); don't (215); like (200); said (189); life (187).
12	*The Witch of Portobello*	Athena (248); said (221); know (207); people (206); love (183); don't (179); I'm (179); time (167); like (158); life (144).
13	*The Winner Stands Alone*	People (351); just (302); know (281); world (261); like (259); time (249); man (241); he's (213); life (199); it's (194)
14	*Aleph*	i'm (220); time (219); know (188); hilal (180); don't (160); people (153); it's (148); say (143); says (143); just (136)
15	*Adultery*	I'm (180); it's (166); like (149); don't (138); love (131); time (125); know (117); life (108); say (108); need (99).

Table 3 represents the ten most frequent words from 15 selected novels which are collected through the Voyant Summary tool. The most frequent words demonstrate the main character and the major themes of the writer. The ten most frequent words from each novel have been examined separately in this research to find out the stylistic fingerprints of Paulo Coelho.

Conclusion

In this study, 15 selected novels have been stylistically analysed through the Voyant Summary tool to find out the stylistic features of Paulo Coelho. The first most common stylistic feature of Paulo Coelho is the use of 'said'; the word 'said' appears in twelve novels from these fifteen novels: 315, 187, 267, 189, 221, 293, 310, 182, 296, 263, and 209 respectively. The use of 'said' indicates that Coelho writes in dialogic and direct narration. The second common feature of Coelho is the use of nouns in his novels as the most frequent words, which reveal the main characters, such as Petrus (360) in *the Pilgrimage*, boy (498) in *the Alchemist*, man (25) in *the Supreme Gift*, hilal (180) in *Aleph*, woman (102) in *by the River Piedra: I Sat Down and Wept*, Paulo (381) in *the Valkyries*, Elijah (420) in *the fifth mountain*, Veronika (195) in *Veronica Decides to Die*, man (120) in *the Devil and Miss Prym*, Maria (231) in *the Eleven Minutes*, I'm (240) in *the Zahir*, Athena (248) in *the Witch of Portobello*, he's (213) in *the Winner Stands Alone*, Brida (367) in the *Brida,* and I'm (180) in *the Adultery*. The third major finding is the themes of Paulo; through the stylistic analysis of these novels. One can validate that love and life are his major themes. As in every novel from these selected novels, he talks about love and life, as love appears for 110, 182, 188, 174, 279, 183, 149, and 149 times, and life appears for 36, 66, 106, 146,

159, 187, 144, 199, 108 times. The fourth major finding is that Paulo does not use difficult and archaic vocabulary, as the ten most frequent words from each novel suggest that it is easy to comprehend by the readers and researchers as well.

This research can be helpful in further studies on the stylistic analysis of any writer or researcher who can continue the research on the same writer. Other future research can be conducted to find out the common stylistic features of other writers through the use of digital humanities tools. This tool use should be integrated into the pedagogical practice of literature.

References

Amancio, D. R. (2015). A Complex Network Approach to Stylometry, 1–21. https:/ doi.org/10.1371/journal.pone.0136076

Eder, M., Rybicki, J., &Kestemont, M. (2016). Stylometry with R: A Package for Computational Text Analysis. The R Journal, 8(1), 107.doi:10.32614/rj-2016-   007

Eder, M., Piasecki, M., &Walkowiak, T. (2017). An open stylometric system based on multilevel text analysis. Cognitive Studies | Étudescognitives, (17). doi:10.11649/cs.1430

Leech, G., & Short, M. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. Great Britain: Pearson Education Ltd.

Manogna, M. L., Charyulu, G. M., & Kumar, G. K. (2018). The quest for the fulfilmentof Destiny: A study on Santiago in Paulo Coelho's "Alchemist.". *International Journal of Computer Engineering in Research Trend*s, p. 5.https://doi.org/10.22362/ijcert/2018/v5/i1/v5i101

Mayer, C., & Maree, D. (2017). A psychobiographical study of intuition in a writer's life: Paulo Coelho revisited. Europe's Journal of Psychology, 13(3), 472–490. doi:10.5964/ejop.v13i3.1184

Mayer, C., & Maree, D. (2018). The Magical Life and Creative Works of Paulo Coelho: A Psychobiographical Investigation. Indo-Pacific Journal of Phenomenology,18(sup1), 1–16. doi:10.1080/20797222.2018.1511310

Modoc, E., &Gârdan, D. (2020). Style at the Scale of the Canon. A Stylometric Analysis of 100 Romanian Novels Published between 1920 and 1940. *Metacritic Journal for Comparative Studies and Theory*, 6(2), 48-63.

O'Sullivan, J., Bazarnik, K., Eder, M., &Rybicki, J. (2018). Measuring Joycean Influences on Flann O'Brien. Digital Studies/Le champ numérique, 8(1). doi:10.16995/dscn.288

Rockwell, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities.* London, England: MIT Press.

Rybicki, J., Eder, M., & Hoover, D. L. (2016). Computational stylistics and text analysis. In *Doing digital humanities* (pp. 159-180). Routledge.

Sarfraz, M. (2022). Stylistic Analysis of Coelho's novel The Alchemist. *International Journal of Linguistics, Literature and Translation*, 5(3), 58–66.

Sinclair, S., & Rockwell, G. (June 29, 2015). Voyant Tools Documentation., University of Alberta, Canada. http://docs.voyant tools.org/

Ullah, Z., Uzair, M., &Mahood, A. (2019). Extraction of Key Motifs as a Preview from 2017 Nobel Prize Winning Novel, 'Never Let Me Go': An Interactive Word Cloud Study.

Ullah, Z., & Mahmood, A. (2019). Stylometry of short stories through voyant corpus summary tool: A text mining study. Kashmir Journal of Language Research, 22(1). 1–17. https://kjlr.pk/index.php/kjlr/article/view/91

Waugh, S., Adams, A., & Tweedie, F. (2000). Computational stylistics using artificial neural networks. *Literary and Linguistic Computing,* 15(2), 187–198.

Wu, K., & Li, D. (2022). Are translated Chinese Wuxia fiction and western heroic literature similar? A stylometric analysis based on stylistic panoramas. *Digital Scholarship in the Humanities*, *37*(4), 1376-1393.