ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

A CORPUS-BASED STYLOMETRIC ANALYSIS OF THE HARRY POTTER SERIES

Alveena Fatima
(MPhil Applied Linguistics scholar) Department of Linguistics and Communications, University of Management and Technology,UMT,Lahore
Email ID: F2024084025@umt.edu.pk
Ms. Shumaila Ahmad
(PhD Linguistics) Department of Linguistics and Communications, University of Management and Technology, UMT, Lahore
Email ID: Shumaila.ahmad@umt.edu.pk
Muhammad Hamzah Masood
(Lecturer) Department of Linguistics and Communications, University of Management and Technology, UMT, Lahore
Email ID: hamzah.masood@umt.edu.pk
Corresponding Author: Muhammad Hamzah Masood
hamzah.masood@umt.edu.pk

ABSTRACT
Text mining technique is mostly involved for the Stylometry of any literary or non-literary text. Previously, Stylometry technique is done manually to find out the written expression of various writers. Being reader it is a tiresome thing to read long texts and find out the words and distinction them as humanely not possible so, the purpose of this study is to extract out the J.K. Rowling's written expressions through text mining technique using Voyant tool as it gives you authentic data (Zafar Ullah, Uzair, & Mahmood, 2019). Moreover, the major objective of this research to reveal the most frequent words of the writer to analyze and distinguish them. Firstly, data was generated as corpus and Summary panel was implied for the stylometric analysis on three alluring series of Harry Potter (1. Harry Potter and the Sorcerer's Stone, 2. Harry Potter and the Chamber of Secrets, 3. Harry Potter and the Goblet of Fire as current study targeted stylistic characteristics qualitatively (interpretation) and quantitatively (statistical data). Mixed method approach was applied using Summary tool. Furthermore, the key-ground theory of Rakesh Aggrawal's Knowledge Discovery Theory (KDD) 1996 (Cabena, Hadjinian, Stadler, Verhees & Zanasi, 1998) which inquires about interesting and new knowledge patterns in dataset as it deals with Stylometry of texts. Major findings in all employs dialogic technique used with "said". The vocabulary density 0.061 to 0.122 and the average length per sentence range12.3 to 14 which express simple and small sentences. The implications help educationist or novice readers to analyse different literal texts for making them concise.

Keywords: *Stylometry, Voyant, corpus, Summary, Text Mining*

1.1. Background Information

Technology, the practical application of scientific knowledge for various purposes, including achieving specific goals, solving problems, giving solutions and improving processes. It involves the use of different types of tools, software, systems, machines, and techniques to operate resources, information and materials.

Technology incorporates the use of different sorts of analytical tools, materials, different techniques, and sources to make life easier or more powerful and productive work in a more pleasant way (Kiyici & Kiyici, 2007). Technology now is a core part of our everyday life. Technology development provides a better quality and comfortable way of life to people (Li & Perkins, 2007).

Technology plays an essential role in the mass production of the corpus to gather the textual or spoken data that shapes the corpus. This involves web scraping tools to scanners for digitizing printed materials, collecting online text; audio recording equipment for spoken corpora etc. Web crawling or Web scraping is a procedure that involves automatic extraction of data that were on different websites using software (Khder, 2021). Corpus technology has been researched down as a rationale research tool in many linguistics and quantitative

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

research for methodology design and identification of linguistic changes. Specialized software tools and libraries designed for corpus linguistics are used to manipulate and analyze corpus data effectively; examples include NLTK (Natural Language Toolkit) and the Corpus Query Language (CQL).

It prevails the generation of multi-modal corpora which includes various communicative modes, such as speech, writing, and spoken. Corpus design is considered the most vital factor of corpus linguistics because it influences the whole analysis and reliability inferences of the corpus *"The design of the corpus is one of the most vital factors in corpus linguistics due to its influence on the whole analysis of the corpus and inferences for the reliability of corpus"* (Rabadi, 2014).

Corpus linguistics helps us to look at what patterns are correlated with lexical and grammatical features. It answers questions like frequent words and phrases used in the text, distinctness between written and spoken texts what type of style occupied by the writer or varieties of tenses used etc. Corpus its plural corpora is a term used in the field of linguistics and shows to a large collection of texts often in electronic forms which is taken as representative of a language and is used to analyse it" (McGillivray, n.d.).

Stylometry or computational stylistics involves the data extraction which shows characteristics of unique writing as Ullah & Mahmood, (2019) said stylistic is like a finger print which is unique character of every person. Text mining is basically a technique which extracts unique and total words, shows vocabulary density, and sentence average length with different themes. This current study use summary tool on harry potter versions for knowledge patterns. Thus, text mining and stylistic qualities together apply on these selected texts.

Joanne Kathleen Rowling, a British author who is famous for Harry Potter fantasy series. These series are most popular because of magical spells. The Harry Potter series comprises of seven books; I applied Stylometry on *"Harry Potter and the Philosopher's Stones, Harry Potter and the Chamber of Secrets and Harry Potter and the Goblet of Fire."*

1.2. Research Objectives

- Find out Summary tool feature represent Stylometry of three versions of Harry potter.
- Find out the stylistic qualities of three versions of Harry potter match or differ with one another?

1.3. Research Questions

1) How does Summary tool feature represent Stylometry of three versions of Harry potter?
2) How do the stylistic qualities of three versions of Harry potter match or differ with one another?

1.4. Research gap

While there have been various studies analyzing different parts of Harry Potter written by J. K. Rowling but conspicuously Stylometry analysis is absent in the existing literature. Existing research leans to focus on stylistic devices or writing patterns manually, often ignoring the broader writing trends and linguistic modulation. A study by Kuzmichenko, A. A. (2023) Provides valuable insights by exploring manually stylistic devices in Harry Potter and the Azkaban's prisoner. Nevertheless, a tool is not used by any researcher for Stylometry analysis of any part of Harry Potter. Moreover, Voyant tool is mostly applied to the literature work rather than to linguistic or any discourse work.

Additionally, this study aims to identify and elucidate the overarching written style that pervade the stylistic differences of Writer, providing a insightful understanding of the discourse dynamics within this diplomatic forum.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

1.5. Statement of the problem

This study is being conducted because reading to long novel is a tiresome thing and many think it's a fool's errand. This tiredness brings difficulty in the analysis of the author written style. "Digital natives" (Prensky, 2001) and technology-obsessed students of the Digi Modern Age want to discover interesting patterns of knowledge and data visualizations with involvement of Digital V&R (Visitor and Resident) model (Ullah, 2019). Using text mining to enhance "digital wisdom" in the embodiment of corpus summaries and collocation patterns to shape dynamic interactive data visual. These stylistic measurements help the readers to easily recognize the writer written style in other written texts too.

In Pakistan, text analytics is still in its inception and the usage of Voyant tools is yet unexercised in the field of linguistic studies and reading methods. So, this work attempts to identify uniform and quantifiable knowledge patterns in Harry Potter and its other versions.

1.6. Significance of the study

The current research explores the most frequent words used in a corpus and these words open the window that Voyant tools are not only used for literature or linguistics studies research but can also be applicable in the field of discourse, cultural, political science and diplomatic relationships of the world. As the study exposes different writing patterns, it is presumed that these themes give concepts about the J. K. Rowling writing style. Ullah Z. & Mahmood, (2019) reveals that stylistics study facilitates elucidation process as stylometric study is important because without it texts cannot comprehend and evaluated in true sense. The study also signifies the choice of words used by writer uncovering the linguistic patterns. The visualization of the words makes the study attractive and interesting for the readers. Therefore, the methodology of this current research removes the element of lethargy from the readers and with less energy they can easily grasp the main concepts or ideas of the largest text corpus without tiresome in minutes.

1.7. Theoretical Modeling

The theory which is used in the current study as a theoretical framework for digitized learning which is Knowledge Discovery Theory in Data Mining (KDD) by Rakesh Aggrawal to gain interesting, intriguing, beneficial and unfamiliar knowledge patterns and learning designs through the evaluation and interpretation of visual data (Cabena, Hadjinian, Stadler, Verhees & Zanasi, 1998). This theory focuses on the variations in the knowledge patterns in the form of statistics, visualization and word form.

The mixed method combines qualitative and quantitative methods (Flick, 2014). It incorporates some disciplines like AI, deep learning, machine learning and hermeneutics (Fayyad, Shapiro, Smyth & Uthurusamy, 1996). The utilization of digital technology innovate the learning process by combining technology, instructional methods, content and pedagogy was called the TPACK Model (Koehler & Mishra, 2009).

1.8. Content/ Manuscript

Here is the list of Harry Potter's parts which are going to analyzed

Table 1*: Manuscript of Three series of Harry Potter*

Series of Harry Potter		
Name of Series	**Author**	**Date**
Harry Potter and the Sorcerer's Stone	Joanne Kathleen Rowling	1st September, 1998
Harry Potter and the Chamber of Secrets	Joanne Kathleen Rowling	2ndJuly , 1998
Harry Potter and the Goblet of Fire	Joanne Kathleen Rowling	8thJuly, 2000

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

1.9. Delimitations

This study was delimited to JK Rowling's only one novel "Harry potter" and its only involving three parts of novel. Further, the current study is limited to the Stylometry analysis of only three most famous series of novel. Moreover, analysis of statistical data is finalized only by using one specific tool of Voyant tool which is Summary tool.

Literature Review

A corpus, in linguistic terms, refers to a massive collection of texts that is used for linguistic analysis and research. These texts can be from various sources, such as books, articles, speeches, or any other form of written or spoken language (Masood et al., 2025). A text corpus comprises of enormous collection of text of billion words built by real language users and used for analyzation of words, phrases and language generic patterns used mostly (Sketch Engine). Several studies were conducted using different corpus based tools and applied to linear and non-linear texts as Wijitsopon (2013) conducted a study using descriptive tools for Keyness, collocates and clusters in Jane Austen's novels which is compared with the BNC corpus. Later on a study was conducted on Robert Frost's poem "Into My Own" for lexical features identification and thematic analysis through the AntConc corpus tool by Saad & Sarbini-Zin (2022) which reflects that Voyant tool is still not in used as Voyant tool is a free-of-charge tool which is a multi-skilled web-based tool for visualization and analysis of texts (Welsh, 2014). The Voyant tools give the readers a view of the canvas of literature for observing the different social themes that were in the text (Asif et al., 2021). Furthermore, Asghar et al. (2021) conducted a study on "waiting for Godot" using UAMTC tool for the existential interpretation of linguistic elements. M. Moustafa (2020) unveiled the thematization of love in Elif' Shafak's The Forty Rules of Love employing the corpus technique which is Word Smith software to generate the keyword list for themes. In contrast to earlier studies, Moustafa (2022) conducted a comparative thematization and characterization study of two South African novels "My Son's Story and Disgrace" that led to the usage of the Sketch engine corpus tool. These studies till now infer that corpus tools applied on different literary novels for various linguistic styles. These researches contributed in my research that these scholars try to documented novelty using different corpus tools except Voyant. Jaafar (2021) analyze one of the significant poems T. S. Eliot's "The Waste Land" through Wmatrix and WebCorp Live corpus tool to examine poetic language keywords and concordances.

In previous years, it is observed that Voyant is considering for literature work as Rambsy (2016) used fictional short stories of Zora Neale Hurston and Richard Wright to find out African American Vernacular English elements as using text mining technique. Ullah & Mahmood (2019) did research on five American short stories to extract quantified stylistic features using a similar tool for vocabulary and sentence construction in all of them. Furthermore, Seif Eddine (2018) used the novel "Bluest Eyes by Toni Morrison" for idealization and self-identification in Pecola's character by identification of common words and their relationships using the Voyant tool. Cirrus tool is also recognized as a word cloud, and embodies a manifestation of Corpus Linguistics, as delineated by Asif et al. (2021). Cirrus as a word cloud shows statistical and visual images of extrusive themes and characters (Ullah et al., 2023). Moreover, Ullah et al., (2019) use a previewing technique for the extraction of themes, key motifs and characters using the cirrus tool on the "Never Let Me Go" novel which opens the gate that different techniques can be evaluated using one tool. Manggong & Rizal (2019) visualize post-colonialism in the "Heart of Darkness" by Joseph Conard through two network-based analyzers Voyant and Cytoscape. Asif et al. (2021) conducted an analysis in which the Voyant technique is used for the identification of

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

characters and themes in Jane Austen's novel Pride and Prejudice as this corpus is already there in this tool. Ullah & Mahmood (2022) researched digital hermeneutic knowledge patterns to explore key themes in Good Bye Mr. Chips by James Hilton using a Voyant text mining tool. Shvetsova et al. (2023) escorted research on the text of P.L. Le Roy's book *"Adventures of four Russian sailors to the island of Spitsbergen brought by a storm, where they lived for six years and three months"* using the same tool.

In (2013) Bonifacio M., a study was on Harry potter using stylistic analysis without implication of any tool for authentic results. Later on, Ullah & Mahmood (2019) implies Voyant tool (Summary tool) on Five American short stories for Stylometry analysis as this gives entire number of documents and words in a corpus. In addition, it provides data on the following topics: unique words; longest and shortest texts based on the number of words in the corpus; highest and lowest vocabulary densities; the average number of words per sentence (highest and lowest); and the five most frequently occurring words in the corpus (Gulati et al., 2023). This research opens a gateway for me. Zhu, H., Lei, L., & Craig, H. (2020) did a Stylometry analysis on Chinese Text "*Dream of the Red Chamber"* but he didn't use any specific tool. Modoc & Gardan (2020) researched on 100 Romanian Novels from 1920 to 1940 to analyze the style using Stylometry technique. (2023) Ahmad & Yaseer used Harry Potter's two parts "*Goblet of Fire and Chamber of secrets*" using Antconc Software for identification of socio-contextual words used by author. All these studies help in finding research gap and open innovative gateways for novice researchers of any field.

This reveals that there is a gap in previous studies which should be filled as any of the researcher worked on most memorable and aesthetic text "Harry Potter" by applying any corpus tool for Stylometry of this specific text.

The previous studies lead to a research gap in linguistics and literature which needs to be filled as several researchers did stylistic analysis study on "Harry Potter series" but none of them incorporate a corpus based tool "Voyant Tool" in their research. Even though they just took one or two Parts of the Novel so, I made corpus of three most amazing parts for better identification and results. I use this "Voyant tool" for authentic data analysis.

3.1 Research Philosophy

The research philosophy which is being incorporated in the current study is positivism as in these we focus on objectivity and measurable data by using summary tool for computationally analyzation of text as it is machine oriented experimental research (Ullah, Z., & Mahmood, A. 2019).

3.2 Research Approach and Paradigm

The research approach which is being incorporated in the current study is mixed-method, consisting of both deductive and inductive approaches. The deductive approach is quantitative and the inductive approach is qualitative research. Schoonenboom (2023) finds out that mixed methods research is commonly combination and integration of quantitative and qualitative data. The current research employs this approach with a basic and general idea of Stylometry of J.K. Rowling's writing style in the Harry Potter parts. The statistics of the corpus analysis are underlying in the quantitative form of research, whereas the interpretation and discretion of the data, fall under the qualitative nature of research.

3.3. Research Design

The research design applied is Non-Experimental as it only deals with the linguistic and stylistic patterns of corpus without manipulation of variables. According to Salmons (2024) it studies different phenomenon without direct manipulation of variables. Along this, it all falls in descriptive Research as it involves the study of stylistic phenomenon such as patterns of word frequency, sentence structure and length without cause and effect relationships. Corpus analysis lies as methodology in descriptive research design.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

3.1.1 Population of the Study

The data of Harry Potter series (three parts) are compiled in Corpus form (corpora) through MS word and then converted into PDF file separately.

3.1.2 Data Sampling

Non-probability sampling was opted for the current research. Non-probability sampling technique is "*totally based on judgments"* (Sharma, 2017). Purposive data sampling technique involves specific chapters that were selected relevant to the research questions. Purposive sampling as the name indicates is applicable when you research for a defined purpose not randomly according to the objectives. It is a sampling method that involves the researcher's personal beliefs and convictions interlinked with study goals and objectives (O & Isaac, 2023).

3.4. Data Collection Tool

The tool which is used for data analysis is Voyant tool "*A text-mining tool*" including Summary panel for statistical stylistic data (e.g., words frequency, sentence length, unique words and vocabulary density). Summary tool automatically gives most frequent words along other details from the corpus. Summary tool gives authentic data as Ullah, Z., & Mahmood, A. (2019) said that summary panel was choose for interpretation of stylistic data.

3.5. Data Collection Procedure

The corpus sampling of three above mentioned series original versions were collected from the official website of the Word Press. After this, copy the text on Microsoft Word and then converted into a PDF file. Three different files were made as it is convenient to upload them easily. Afterwards, uploaded the PDF file on the Voyant tool and analyzed the results using the Summary tool. The generated data included into discussion section of research paper. After this, statistical data were analysed for stylistic patterns of author's writing style. Text data's clarification was called *"a bag of words"* (Aggarwal & Zhai, 2012. P. 3).

3.6. Nature of the Data

The analyzed data was primary in nature and obtained from official and authentic website of Harry Potter novel as according to

3.7. Data Analysis Method

Voyant tools were applicative in this research and it is a web-based text analyzation and reading software that facilitate users to read and explore different types of texts using a multi-functional interface. It was created by Stefan Sinclair of McGill University and Geoffrey Rockwell (2003) of the University of Alberta for digital humanities students and scholars, the general public, and those new to digital humanities. It is an open-source and free software package that lays out multifaceted and advanced text analysis for beginners and advanced ones. It is a combination of 51 different tools that helps the users to see the most chronic words along the context and their frequencies of the document or text which they have chosen. Thereupon, this is further categorized into various data visualization forms. These tools of Voyant tools deeply visualize the findings and results of quantitative analysis of qualitative data (Hetenyi, Lengyel, & Szilasi, 2019). It gives the user options to select a ready-made corpus or own corpus. Moreover, you can use the different URL's corpus too.

The theory which is used in the current study as a theoretical framework for digitized learning which is Knowledge Discovery Theory in Data Mining (KDD) by Rakesh Aggrawal to gain intriguing, beneficial, interesting, unfamiliar patterns of knowledge and learning designs through the evaluation and interpretation of visual data (Cabena, Hadjinian, Stadler, Verhees & Zanasi, 1998). This theory focuses on the variations in the knowledge patterns in the form of statistics, visualization and word form. Using this theory as a base, in this research different writing pattern of the author was observed with the help of numerical data that was extracted with the help of this tool.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

The research procedure was that Word file of Harry Potter's of three parts were first created through official website for authentic data and compiled in different Microsoft Word files. After compilation, converted into PDF files and later on those files were separately uploaded on the Voyant tool and stripped file data in the five panels to make it plain text and remove extra data like chapters name, page numbers or images. The current study is centralized on the Summary tool so; most frequent words along with frequency were shown. Later, it was emulate and then evaluated its statistical and visual data.

The statistical data shows different characters name and some verbs which identify the author's writing style. The corpus was uploaded three times to check the accuracy of tool and every time the same result was displayed. The files were uploaded one by one for data extraction and to easily compare and observe the author style easily without a mess with other files.

3.8. Strategy of Data Analysis

The research strategy of data analysis is done by using a tool "Summary Tool." as when we upload a corpus file, it displays different portions of different tools and there was a one portion in which we can easily find the statistical data of our corpus with detailed information. That numerical data was analysed for the Stylometry of text by reading them at one click in the text summary.

Manually, it was a tough task for the researcher to identify stylistic features of each part separately. Furthermore, it was a time consuming to count each word and separating unique vocabulary along with sentence structures as there might chances of human errors. It's impossible to count average words number in each gigantic part as this required a long span. It's a fool errand to count words and make a list. Therefore, Voyant tool gives us different statistical values at one click and it works like a magic.

4.1 Results and Discussion

In this section, Summary tool result is pasted as in figure form and stylistic characteristics revealed novelty knowledge patterns in the uploaded corpus. Furthermore, knowledge patterns are extracted as most frequent words in these selected series. Statistical data is revealed through Summary panel. In summary tool the five most frequently occurring words have shown (Gulati et al., 2023). In the following lines, stylistic characteristics of three series are interpreted.

Harry Potter and the Sorcerer's Stone by J.K. Rowling

This corpus has 1 document with 75,190 total words and 8,270 unique word forms. Created about a minute ago.

Vocabulary Density: 0.110

Readability Index: 82.966

Readability Index: 82.966

Average Words Per Sentence: 12.3

Most frequent words in the corpus:

- harry (1181); said (779); ron (399); hagrid (325); hermione (237); know (202); got (186); didn't (178); like (174); just (171); professor (163); looked (151); it's (147); don't (137); snape (132); he'd (129); dumbledore (129); going (126); right (121); look (121); i'm (119); uncle (114); think (114); time (113); potter (112); dudley (106); vernon (101); malfoy (101); looking (101); harry's (100); eyes (98); door (97); neville (95); couldn't (95); yeh (94); stone (93); head (90); come (90); told (89); face (88); thought (87); mcgonagall (87); way (86); good (85); quirrell (80); people (80); that's (79); room (79); boy (78); ter (77); left (75); hogwarts (75); gryffindor (75); he's (73); mr (72); i've (72); heard (71); wasn't (70); turned (70)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.8.No.2 2025

Figure 1: *Statistical Data of Harry Potter and the Sorcerer's Stone*

Stylometry, a technique quantifies stylistic characteristics of any literary work. Basically, Joanne Kathleen Rowling, Harry Potter's Series writer, writes 8,270 unique words and he repeats these more than nine times until they appear 75, 190 total words. This repetition creates ease for literature lovers to memorize the story and can easily connect events. The vocabulary density is 0.110 which is quite easy for novice readers to understand about fiction, description, locations and it provide continuity of scenes. According to Ullah, Uzair & Mehmood, (2019) vocabulary density is taken out by division of unique words and total words. Readability quotient is 82.966 which show high reading difficulty level and person with college-intermediate knowledge of vocabulary can easily understand it. Average words per sentence lies at 12.33 which unveil the use of long sentences (compound and complex sentences) aiding readability for readers. Other stylistic aspect of small sentences unveil that simple sentences make story comprehensive for all age readers. Moreover, this part is a fantasy fictional in nature and use of small, simple sentences along unique magical words make her style different and friendly for all age group readers.

Statistical data shown in figure reveals knowledge patterns of major characters of the harry Potter for example "Harry (1181)", "Ron (399)," "Hagrid (325)'" and "Hermione (237)," are the first five most occurring characters that driven the narration and storyline. The repetition shows their importance in story. Furthermore, "Harry's (100)," "eyes (98)," other most repeated words serves a reminder of emotional depth between characters. Apart from these, other most non-living character of the story "Stone (93)," occur which refers to the title of this series as it plays a magnificent role. This stone was a magical one and attached with Narrator's life.

Other stylistic feature is "said (77)," occur as second most frequent word which unveils dialogic style of the writer. Along this "know (202)," and "got (186)," reveals substantial dialogue interaction between Harry, Hagrid, Ron, Hermione, Malfoy, professor Snape, Dumbledore and Mcgonagall as most prominent characters that lead story to resolution. Additionally, use of "didn't (178)," "don't (132)," and "looking (101)," shows conversational tone. Besides this frequent use of names like "Hogwarts (75)," and "Gryffindor (75)," manifest Harry Potter magical universe. Moreover, "I'm (119)" suggests first person narrative technique as in dialogues it is the most reliable source of narration. These techniques distinguish J.K. Rowling's stylistics.

Harry Potter and the Chambers of the Secrets by J.K. Rowling

This corpus has 1 document with 82,442 total words and 10,027 unique word forms. Created 25 seconds ago.

Vocabulary Density: 0.122

Readability Index: 9.951

Average Words Per Sentence: 14.0

Most frequent words in the corpus:

- harry (1382); said (1181); ron (621); hermione (261); malfoy (189); got (178); professor (176); lockhart (175); just (171); like (169); know (162); it's (145); weasley (143); didn't (141); time (139); don't (139); looked (137); think (136); eyes (131); harry's (130); dobby (129); right (126); mr (121); looking (121); face (121); potter (120); dumbledore (119); hagrid (115); look (114); door (114); head (111); come (109); long (108); voice (103); i'm (103); going (103); wand (99); riddle (98); school (97); fred (97); way (90); room (90); ginny (89); away (87); sir (86); told (85); behind (84); large (83); hogwarts (83); heard (83); mrs (82); dark (81); car (81); oh (79); hand (79); chamber (79); tell (78); george (78); slytherin (77)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.8.No.2 2025

Figure 2: *Statistical Data of Harry Potter and the Chamber of Secrets*

In this second figure, 10,027 unique words were written by writer with total words 82,442. So its vocabulary density is 0.122 and it shows richer words appear in his text. As compared to first, it has more advanced vocabulary addition. The average word per sentence is 14.0 which reveal that more complex sentences are structured in written expressions. Furthermore, readability index is 82.99 lowered than previous suggests it's an easier corpus and accessible to wide range of audiences.

Involvement of every character led to the J.K.'s new innovative style as "Harry (1382"), "Ron (621)," "Hermione (261") depict these are the prominent figures and shows centrality of text. They gone through dialogic style as "said (1181)," "think (136)," and "voice (103)" used for conversation and dialogue delivery which focus on interpersonal interactions. Moreover, "Chamber (79)" the non-living character repeated to show the emphasis and its importance as it's too in the title. Other important non-living character is "wand (99)" which shows connection of magic with characters and makes the series alluring and eye-catchy. Writer uses new Charactonyms like "Weasely (143)," "Hogwarts (83)," and "Slytherin (77)," shows the novice writing style.

Additionally, words like "Wand (99)," "riddle (98)," "school (97)," "room (90)" and "car (81)" emphasize about magical setting and academic environment picturize by the writer. Lexical like "large (83)," and "dark (81)," demonstrates magical theme of writer's writing style. Harry Potter and the Chambers of the Secrets by J.K. Rowling

This corpus has 1 document with 197,808 total words and 12,123 unique word forms. Created 13 seconds ago.

Vocabulary Density: 0.061

Readability Index: 7.609

Average Words Per Sentence: 13.9

Most frequent words in the corpus:

- harry (2933); said (2638); ron (979); hermione (807); mr (538); looked (489); dumbledore (488); like (448); just (438); got (418); know (386); weasley (370); chapter (353); looking (345); hagrid (320); moody (317); crouch (317); eyes (316); right (308); didn't (306); face (300); don't (295); look (280); time (278); professor (278); wand (276); going (270); come (261); voice (253); saw (238); harry's (238); potter (236); head (236); it's (233); think (229); ing (224); cedric (224); told (218); bagman (214); dark (212); room (209); snape (207); thought (206); way (203); i'm (198); he's (193); turned (191); krum (190); long (187); hogwarts (187); sirius (186); fred (186); house (185); hand (184); heard (180); away (179); good (178); table (176); great (174)

Figure 3: *Harry Potter and the Goblet of Fire*

In the above statistical data, Stylometry of J.K. Rowling is shown as this corpus comprises of 197,808 total words having 12,123 unique word forms which portrays rich lexicon diversity of words. Repetition shows thematic consistency. Vocabulary density is 0.061 which is lowest rather than other. This lower density shows tendency of repeated terms for proper focus of readers. Furthermore, Average words per sentence are 13.9 which signal that moderately and descriptive narration is involved. Readability index is 7.609 aligns with intermediate readers.

As protagonist is the one who leads the story. "Harry (2933)" repeated frequently to engage the reader's attention. Moreover, other famous and main side characters are repeated too as "Ron (979)," "Hermione (807)," "Dumbledore (488)," "Hagrid (320)," "Cedric (224)," "Krum (190)," "Sirius (186)," and "Fred (186)" shows interactions and relations. Conversational style is still intricate in this as verbs like "said (2638)," recurrent as second most highest in the corpus. Other dialogic words "told (218)," "thought (206)," proves narration style.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Quantitative analysis of three harry Potter series have been presented in a tabular form for holistic overview

Table 1 *Quantitative Corpus of Three Series*

	Harry Potter series	Total Words	Unique Words	Vocabulary Density	Length of sentences
Series 1	**Harry Potter and the Sorcerer's Stone**	75,190	8,270	0.110	12.3
Series 2	**Harry Potter and the Chamber of Secrets**	82,442	10,027	0.122	14.0
Series 3	**Harry Potter and the Goblet of Fire**	197,808	12,123	0.061	13.9

The table 1 reveals that vocabulary density increases and then decreases ranges from 0.110 to 0.061 which is relatable for novice readers. Increment in total and unique words reveals that suitable for every reader to build up new words bank easily. The frequent word usage make connections for novice reader to build interconnectivity of words enhance readability level. The average length ranges from 12.3 to 14.0 shows appropriateness for novel readers.

Qualitative data of study presented in table 2

Table 2 *Qualitative Data of Three Harry Potter series*

	Three Harry Potter Series	*Qualitative Data*
Series 1	**Harry Potter and the Sorcerer's Stone**	harry (1181); said (779); ron (399); hagrid (325); hermione (237)
Series 2	**Harry Potter and the Chamber of Secrets**	harry (1382); said (1181); ron (621); hermione (261); malfoy (189)
Series 3	**Harry Potter and the Goblet of Fire**	harry (2933); said (2638); ron (979); hermione (807); mr (538)

Text mining technique shows the Stylometry of writer as the most frequent characters name, themes, narrated verbs repeated to show the connection to novice readers.

5.1 Key findings

This study focused on quantified stylistic qualities of three Harry Potter series. First most prominent stylistic features are characteronym which unveils key characters of the series and

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

their repetition shadows on their importance in the story as "Harry, Hermione, Ron, and Dumbledore" are the key leading role. Frequent use of characters names differentiate J.K. Rowling's style from others and make it prominent. Furthermore, the second most focused outcome is "said" occurs as 779, 1181, and 2683 as it increases and gives a glimpse of dialogical narration style used by writer. Third key finding is words like "dark (81) (212), "voice (253) (103)," shows the black, magical theme of all series and writer is famous for such writing style which makes her work distinguished from others. Fourth key finding is that all these selected series have wide range of vocabulary which increases double times in each new part that accelerates reading speed provide ease to readers. Last finding is that average sentence per length enhances readability index for all age group readers.

5.2 Implications of the study

The study is of interest to academic researchers to research on different literary texts of their choice to check their impact on learners or readers psychologically; as this implication is helpful for different primary, medium and advanced level readers and learners. Furthermore, short stories of primary and intermediate short stories, subjective essays, assignments, translated texts can be evaluated accurately and precisely. This study methodology can also apply on the political speeches for thematic analysis using Voyant tool other tools as cirrus tool. Moreover, in forensic linguistic, suicidal notes can be evaluated using summary tool that in which context words used by them. It's not only applicable to linguistic or literature field but also to other domains of social sciences, engineering etc.

5.3 Future recommendations

The findings of this research offer valuable insights into the three specific series of renowned novel through text mining technique for identification of Stylometry of writer. Furthermore, this research can be applied by researchers to do comparison between two writers' text and style. Additionally, to analyse short stories of Punjab textbooks for repetitive characters and themes. Novice educators follow their written footprints to earn nobility (Roohi et al., 2021). Besides, objectionable political discourse can be thematically analysed using summary tool as futuristic approach.

References

(Boni) Li, Y., & Perkins, A. (2007). The impact of technological developments on the Daily life of the elderly. *Technology in Society*, *29*(3), 361-368. https://doi.org/10.1016/j.techsoc.2007.04.004

Agar, J. (2019). What is technology? *Annals of Science*, *77*(3), 377-382. https://doi.org/10.1080/00033790.2019.1672788

Asghar, I., Abbas, S., Bashir, M., & Ahmad, N. N. (2021). Investigating existential interpretation of waiting for godot: A corpus-based study of lexical features. *Humanities & Social Sciences Reviews*, *9*(3), 1314-1324. https://doi.org/10.18510/hssr.2021.93130

Asif, M., Shafqat, S., & Yasmeen, S. (2021). A Corpus-Based Study of Austen's Novel Pride & Prejudice using Cirrus Tool. *Corporum: Journal of Corpus Linguistics,*, *4*(1).

BR Web Desk. (2019, September 28). *Full transcript of prime minister Imran Khan's speech at the UNGA*. Brecorder. https://www.brecorder.com/news/524851

Carstens, K. J., Mallon, J. M., Bataineh, M., & Al-Bataineh, A. (2021). Effects of technology on student learning. *The Turkish Online Journal of Educational Technology*, *20*(1). https://files.eric.ed.gov/fulltext/EJ1290791.pdf

Chen, X., Yan, Y., & Hu, J. (2019). A corpus-based study of Hillary Clinton's and Donald Trump's linguistic styles. *International Journal of English Linguistics*, *9*(3), 13. https://doi.org/10.5539/ijel.v9n3p13

Dash, N. S. (2008). *Corpus linguistics: An introduction*. Pearson Education India.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Davurovna, M. M. (2019). Usage of IT terminology in corpus linguistics. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(7). https://www.ijitee.org/wp-content/uploads/papers/v8i7s/G10390587S19.pdf

Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-methods research: A discussion on its types, challenges, and criticisms. *Journal of Practical Studies in Education*, *2*(2), 25-36. https://doi.org/10.46809/jpse.v2i2.20

Full text of PM Modi's address at 75th UNGA session 2020. (2020, September 26). Hindustan Times. https://www.hindustantimes.com/india-news/full-text-of-pm-modi-s-address-at-75th-unga-session-2020/story-8sOpyLIDesleUtIhRYdIWJ.html

Full text of PM Narendra Modi's UNGA speech. (2019, September 28). India Today. https://www.indiatoday.in/india/story/pm-narendra-modi-un-unga-full-speech-text-new-york-us-1604079-2019-09-28

FutureLearn. (2022, October 25). *What is a corpus?* https://www.futurelearn.com/info/courses/understanding-dictionaries/0/steps/62809

Granberg, E. M. (2000). *How technology enhances teaching and learning*. Vanderbilt University. https://cft.vanderbilt.edu/articles-and-essays/the-teaching-forum/how-technology-enhances-teaching-and-learning/

Grover, V. (2015). RESEARCH APPROACH: AN OVERVIEW. *International Multidisciplinary Research Journal GRT*, *4*(8). https://www.researchgate.net/publication/273352276_RESEARCH_APPROACH_AN_OVERVIEW

Gulati, S., Sinhababu, A., & Chakravarty, R. (2023). Understanding the research landscape of smart libraries using text mining and data visualization : A use case of Voyant tool. *COLLNET Journal of Scientometrics and Information Management*, *17*(1), 159-174. https://doi.org/10.47974/cjsim-2022-0051

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, *3*, 275-285. https://doi.org/10.1016/j.susoc.2022.05.004

Hetenyi, G., Lengyel, A., & Szilasi, M. (2019). Quantitative analysis of qualitative data: Using voyant tools to investigate the sales-marketing interface. *Journal of Industrial Engineering and Management*, *12*(3), 393. https://doi.org/10.3926/jiem.2929

India's statement delivered by the external affairs minister, Dr. S. Jaishankar at the general debate of the 77th session of the UN General Assembly. (2022, September 25). Ministry of External Affairs, Government of India. https://www.mea.gov.in/Speeches-Statements.htm?dtl/35795/

InfoGuides: Text analysis tools: Voyant. (2023, May 18). InfoGuides at George Mason University. https://infoguides.gmu.edu/textanalysistools/voyant#

Jaafar, E. A. (2021). Poetic language and the computer: A corpus stylistic study of the waste land. *International Journal of Language and Literary Studies*, *3*(1), 230-239. https://doi.org/10.36892/ijlls.v3i1.525

Khder, M. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications*, *13*(3), 145-168. https://doi.org/10.15849/ijasca.211128.11

KIYICI, F. B., & KIYICI, M. (2007). SCIENCE, TECHNOLOGY & LITERACY. *The Turkish Online Journal of Educational Technology – TOJET*, *6*(2). https://files.eric.ed.gov/fulltext/ED500046.pdf

Kuzmichenko, A. A. (2023). "HARRY POTTER AND THE PRISONER OF AZKABAN": OVERVIEW OF STYLISTIC DEVICES. *Russian Linguistic Bulletin*, (12 (48)), 13.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Moustafa, B. S. (2022). Thematization processes and the representation of love in the forty rules of love: A corpus stylistic analysis. *Eurasian Journal of Applied Linguistics, 8*(1). https://ejal.info/menuscript/index.php/ejal/article/view/195

Manggong, L., & Rizal, M. (2019). Postcolonial network analysis of Joseph Conrad's heart of darkness. *Proceedings of the Proceedings of the 3rd English Language and Literature International Conference, ELLiC, 27th April 2019, Semarang, Indonesia*. https://doi.org/10.4108/eai.27-4-2019.2285320

Masood, M. H., Saher, A., & Yasir, H. S. (2025). Students attitudes toward learning English as a second language in colleges of Gilgit-Baltistan. *Qlantic Journal of Social Sciences*, *6*(1), 252-268. https://doi.org/10.55737/qjss.vi-i.25301

Mithe, R., Indalkar, S., & Divekar, I. (2013). Optical Character Recognition. *International Journal of Recent Technology and Engineering (IJRTE), 2*(1). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6a4b4f04d5ce3c3592832eb40c23cc8fc5a9131e

Moustafa, B. S. (2022). A comparative corpus stylistic analysis of thematization and characterization in gordimer's *My son's Story* and coetzee's *Disgrace*. *Open Linguistics, 8*(1), 46-64. https://doi.org/10.1515/opli-2020-0183

NDTV news. (2023, September 26). "There's Much India Has To Share". *NDTV*. https://www.ndtv.com/india-news/theres-much-india-has-to-share-full-text-of-s-jaishankars-speech-at-un-4426133

Nightingale, P. (2014). What is technology? Six definitions and two pathologies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2743113

O, & Isaac, E. (2023). Convenience and Purposive Sampling Techniques: Are they the Same? *International Journal of Innovative Social & Science Education Research, 11*(1), 1-7. https://seahipaj.org/journals-ci/mar-2023/IJISSER/full/IJISSER-M-1-2023.pdf

ÖZBAY, A. Ş., & GÜRSOY, Z. (2023). Computerized corpus as a tool for educational technology and learning in the analysis of four-word recurrent expressions. *Journal of Educational Technology and Online Learning*, *6*(1), 249-272. https://doi.org/10.31681/jetol.1186346

Ozone, S., Haruta, J., Takayashiki, A., Maeno, T., & Maeno, T. (2020). Students' understanding of social determinants of health in a community-based curriculum: A general inductive approach for qualitative data analysis. *BMC Medical Education, 20*(1). https://doi.org/10.1186/s12909-020-02391-z

Pole, K. (2007). Mixed Method Designs: A Review of Strategies for Blending Quantitative and Qualitative Methodologies. *Mid-Western Educational Researcher, 20*(4). https://d1wqtxts1xzle7.cloudfront.net/41630452/Mixed_Methods_Designs_A_Review_of_Strate20160127-7390-9muqfs-libre.pdf

Prime minister Imran Khan's address at 75th United Nations General Assembly session [Video+Transcript]. (2020, September 25). Pakistan Tehreek-e-Insaf. https://www.insaf.pk/news/pm-ik-address-75th-unga

Rabadi, R. I. (2014). The Use of Corpora in Linguistics. https://www.researchgate.net/publication/290834364_The_Use_of_Corpora_in_Linguistics

Rambsy, K. (2016). Text-Mining Short Fiction by Zora Neale Hurston and Richard Wright using Voyant Tools. *College Language Association, 59*(3), 251-258. //www.jstor.org/stable/44325917

Rayson, P., Walkerdine, J., Fletcher, W. H., & Kilgarriff, A. (2006). Annotated web as corpus. *Proceedings of the 2nd International Workshop on Web as Corpus - WAC '06*. https://doi.org/10.3115/1628297.1628302

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Research Design: What It Is (Plus 20 Types). (2022, August 8). https://www.indeed.com/career-advice/career-development/research-design

Roohi, T., Masood, M. H., & Mushtaq, R. (2021). An analysis of equity in secondary school education: A Headteachers' perspective. *Global Educational Studies Review, VI*(III), 29-40. https://doi.org/10.31703/gesr.2021(vi-iii).04

Saad, A. H., & Sarbini-Zin, M. (2022). Corpus-Assisted Analysis of Robert Frost's Poem, "Into My Own" Using AntConc. *PENDETA*, *13*(1), 1-9. https://doi.org/10.37134/pendeta.vol13.1.1.2022

Schoonenboom, J. (2023). The Fundamental Difference Between Qualitative and Quantitative Data in Mixed Methods Research. *FQS*, *24*(1). https://www.qualitative-research.net/index.php/fqs/article/view/3986/4917

Schumann, F. R. (2022). Using Voyant Tools for Data Mining Social Media Comments about a Destination: A Guam Study. *Journal of Hospitality & Tourism*, *20*(2). https://johat.org/wp-content/uploads/2022/12/6.pdf

Seif Eddine, N. A. (2018). The Idealization and Self-Identification of Black Characters in the Bluest Eyes by Toni Morrison: Using Voyant Text Analysis Tools. *Journal of Literature, Languages and Linguistics*, *49*. https://core.ac.uk/download/pdf/234693608.pdf

Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research*, *3*(7), 749-752. https://d1wqtxts1xzle7.cloudfront.net/58765080/Pros_and_cons_of_sampling-libre.pdf

Shvetsova, T., Dulova, S., & Shakhova, V. (2023). Voyant tools as an analysis instrument of Pierre-Louis Le Roy's text. *E3S Web of Conferences*, *420*, 06008. https://doi.org/10.1051/e3sconf/202342006008

Shvetsova, T., Dulova, S., & Shakhova, V. (2023). Voyant tools as an analysis instrument of Pierre-Louis Le Roy's text. *E3S Web of Conferences*, *420*, 06008. https://doi.org/10.1051/e3sconf/202342006008

Trochim, W. M. (2004). *Research methods: The concise knowledge base*.

Ullah, Z. (2019). Unveiling Knowledge Patterns in Intermediate English Textbooks Through Voyant Text Mining Tools: A Digital Humanities Study. *National University of Modern Languages, Islamabad*.

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). http://kjlr.pk/index.php/kjlr/article/view/91/64

Ullah, Z., & Mahmood, A. (2022). Voyanting hermeneutic patterns from James Hilton's Novel Goodbye, Mr. Chips: A text mining study. *Webology*, *19*(2). https://orcid.org/0000-0003-3773-1467

Ullah, Z., Rafique, M., & Arshad, R. (2023). Tracing Intersectionality, Individual and Familial Struggle in Morrison's Novels: A Digital Humanities Study. *Corporum: Journal of Corpus Linguistics*, *6*(1).

Ullah, Z., Uzair, M., & Mahmood, A. (2019). Extraction of key motifs as a preview from 2017 Nobel Prize Winning Novel, 'Never Let Me Go': An Interactive Word Cloud Study. *ISSN 2306-112X (E)*, *7*(2), 2305-6533. https://d1wqtxts1xzle7.cloudfront.net/85034495/28-libre.pdf?1651046187=&response-content-disposition=inline%3B+filename%3DExtraction_of_Key_Motifs_as_a_Preview_fr.pdf&Expires=1701540272&Signature=Zrb2xcfuWlhucV6NtCqMdNJrz-wKEN8Ukf5OZnuUjT5vVj4xnJtn-

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

ZhyLC2wlRZ~5xYIBfa1hJya3ZOJwk~zLTCO5YBoUaRHREZT2qAHj5YIGOwXww63GBxoKhfMSsNSrcWKmLlYhV2faPnBNcbccQt72MR~3XeFAwOZnPj22kJazssCrkBY9J0NXoc5PXl6uqmDPCog6LMvJn6p7bUWGVn-3YIiv9jNgIN8LJZbdOT5~hvS76hjkWX7N1Jy1i6JmwLr9iDyKP7i4uRTlmtMO2BE1FlobApYuK~l~F15KgcgZuTWKR9Y8-K7HgpwoCshLvV87GlHdpCjEUfo8vLJ5Q__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Wagner, S., & Marshall, A. (2019). Talking About Research: Applying Textual Analysis Software To Student Interviews. *ACRL*. https://alair.ala.org/bitstream/handle/11213/17688/TalkingAboutResearch.pdf?sequence=1

Watson, R. (2015). Quantitative research. *Nursing standard*, *29*(31), 44-48. https://doi.org/10.7748/ns.29.31.44.e8681

Web Desk. (2021, September 25). *Full transcript: PM Imran Khan's speech at 76th session of United Nations General Assembly*. ARY NEWS. https://arynews.tv/pm-imran-unga-speech-76-session/

Welsh, M. E. (2014). Review of Voyant Tools. *Collaborative librarianship*, *6*(2). https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1105&context=collaborativelibrarianship

What is speech in language studies? (2009, June 10). ThoughtCo. https://www.thoughtco.com/speech-linguistics-1692121#

When India grows, the world grows, when India reforms, the world transforms: PM Modi. (2021, September 25). www.narendramodi.in. https://www.narendramodi.in/text-of-prime-minister-narendra-modi-s-address-at-unga-557522

Wijitsopon, R. (2013). A corpus-based study of the style in Jane Austen's novels. *MANUSYA*, *16*(1), 41-64. https://doi.org/10.1163/26659077-01601003

Yılmaz, E., & Soruç, A. (2015). The use of concordance for teaching vocabulary: A data-driven learning approach. *Procedia - Social and Behavioral Sciences*, *191*, 2626-2630. https://doi.org/10.1016/j.sbspro.2015.04.400

Yoon, H. (2008). MORE THAN A LINGUISTIC REFERENCE: THE INFLUENCE OF CORPUS TECHNOLOGY ON L2 ACADEMIC WRITING. *Language Learning & Technology*, *12*, 31-48. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/9d7d4802-8866-4246-af19-69187c5f14bc/content

Zaidi, S., & Allahdad, S. (2023). Analyzing Recent (2019) Kashmir Socio-Political Issue: A Voyant Sentiment Analysis of Tweets. *Linguistics and Literature Review (LLR)*, *9*(2). https://doi.org/10.32350/llr.92.02

Zaidi, S., & Allahdad, S. (2023). Analyzing Recent (2019) Kashmir Socio-Political Issue: A Voyant Sentiment Analysis of Tweets. *Linguistics and Literature Review (LLR)*, *9*(2). https://journals.umt.edu.pk/index.php/llr/article/view/4018/1918

Zompetti, J. P. (2019). Rhetorical Incivility in the Twittersphere: A Comparative Thematic Analysis of Clinton and Trump's Tweets During and After the 2016 Presidential Election. *Journal of Contemporary Rhetoric*, *9*, 29-54. https://www.researchgate.net/profile/Joseph-Zompetti/publication/338189571_Rhetorical_Incivility_in_the_Twittersphere_A_Comparative_Thematic_Analysis_of_Clinton_and_Trump's_Tweets_During_and_After_the_2016_Presidential_Election/links/5e05f450a6fdcc283742194a/Rhetorical-Incivility-in-the-Twittersphere-A-Comparative-Thematic-Analysis-of-Clinton-and-Trumps-Tweets-During-and-After-the-2016-Presidential-Election.pdf