

**THEMATIC INVESTIGATION OF *SORROWS*  
*OF SARASVATI*: A VOYANT TOOLS TEXT  
MINING APPROACH IN DIGITAL  
HUMANITIES**

**BY**

**SHAHZEB SHAFI**



**NATIONAL UNIVERSITY OF MODERN LANGUAGES  
ISLAMABAD**

**FEBRUARY, 2024**

**Thematic Investigation of *Sorrows of Sarasvati*: A Voyant  
Tools Text Mining Approach in Digital Humanities**

By

**SHAHZEB SHAFI**

MA Education, Sarhad University of Science and Information Technology, Peshawar

2023

BS English Language, National University of Modern Languages, Islamabad, 2020

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

**MASTER OF PHILOSOPHY**

In **ENGLISH**

To

FACULTY OF ARTS & HUMANITIES



NATIONAL UNIVERSITY OF MODERN LANGUAGES, ISLAMABAD

© Shahzeb Shafi, 2024



## THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with the overall exam performance, and recommend the thesis to the Faculty of Arts & Humanities for acceptance.

**Thesis Title:** Thematic Investigation of Sorrows of Sarasvati: A Voyant Tools Text Mining Approach in Digital Humanities

**Submitted by:** Shahzeb Shafi

**Registration #:** 67-MPhil/Eng-Ling/F20

Master of Philosophy

Degree name in full

English Linguistics

Name of Discipline

Dr. Arshad Ali

Name of Research Supervisor

\_\_\_\_\_  
Signature of Research Supervisor

Prof. Dr. Muhammad Safeer Awan

Name of Dean (FAH)

\_\_\_\_\_  
Signature of Dean (FAH)

\_\_\_\_\_  
Date

## CANDIDATE DECLARATION FORM

I **Shahzeb Shafi**

Son of **Shafi Muhammad**

Registration # **67-MPhil/Eng-Ling/F20**

Discipline **English Linguistics**

Candidate of **Master of Philosophy** at the National University of Modern Languages do hereby declare that the thesis **Thematic Investigation of Sorrows of Sarasvati: A Voyant Tools Text Mining Approach in Digital Humanities** submitted by me in partial fulfilment of MPhil degree, is my original work, and has not been submitted or published earlier. I also solemnly declare that it shall not, in the future, be submitted by me for obtaining any other degree from this or any other university or institution.

I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be cancelled and the degree revoked.

\_\_\_\_\_  
Signature of Candidate

**Shahzeb Shafi**  
\_\_\_\_\_  
Name of Candidate

\_\_\_\_\_  
Date

## ABSTRACT

### **Title: Thematic Investigation of *Sorrows of Sarasvati*: A Voyant Tools Text Mining Approach in Digital Humanities**

This research employed a text mining approach using five (5) computational Voyant Tools, namely the Cirrus tool, Summary Tool, Contexts tool, Trends tool, and Collocates tool, to explore different themes from the translated novel *Sorrows of Sarasvati: The Lost River*. The novel is a translation of a famous Pakistani writer Mustansar Hussain Tarar's novel *Bahao* by Dr. Muhammad Safeer Awan and Dr. Saleem Khan, published in 2021. The research onion of Saunders, Lewis, and Thornhill (2012) was employed as a research methodology to explore novel themes through distant reading. The quantitative results of the study are presented through text mining which consisted of data visualization (Cirrus), the most occurring terms with their statistical frequencies (Summary tool), collocations of the most frequent words (Collocates tool), and graphs of these terms and collocations in ten (10) distinct segments of the novel (Trends tool); whereas the qualitative data constituted of the themes obtained through the analysis of tools mentioned above along with the contexts of most frequent words (Contexts Tool). The findings of text mining of Corpus of *Sorrows of Sarasvati: The Lost River* revealed data visualization, the most frequent terms with their frequencies, trends and collocations, and nomenclature; while the thematic analysis extracted the themes of nature and wilderness, mystery and unknown, interplay between existence and natural environment, death, loss and sorrow, emotions sensations and nostalgia, transformations in nature and environmental changes, impermanence and transience, journey, human fragility, communication and knowledge, community and shared responsibility, and birds' anatomy.

**Keywords:** *Sorrows of Sarasvati: The Lost River*, Voyant Tools, Text Mining, Thematic Analysis, Digital Humanities.

## TABLE OF CONTENTS

<b>THESIS AND DEFENSE APPROVAL FORM.....</b>	<b>ii</b>
<b>CANDIDATE DECLARATION FORM.....</b>	<b>iii</b>
<b>ABSTRACT .....</b>	<b>iv</b>
<b>TABLE OF CONTENTS.....</b>	<b>v</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>viii</b>
<b>LIST OF FIGURES .....</b>	<b>ix</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>xi</b>
<b>DEDICATION .....</b>	<b>xii</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 Statement of the Problem .....	3
1.2 Research Objectives .....	5
1.3 Research Questions .....	6
1.4 Significance and Rationale of the Study .....	6
<b>2. LITERATURE REVIEW .....</b>	<b>9</b>
2.1 Introduction .....	9
2.2 Conceptual Literature .....	9
2.2.1 Text Mining .....	9
2.2.2 Corpus .....	10
2.2.3 CL and DH .....	10
2.2.4 History of DH .....	11
2.3 Voyant Tools .....	12
2.3.1 Cirrus tool .....	13
2.3.2 Summary tool .....	13
2.3.3 Contexts tool .....	14
2.3.4 Trends tool .....	15
2.3.5 Collocates tool .....	15
2.4 Research Literature on VT .....	15
2.5 Literature on OMT .....	21
2.6 Limitations .....	23

2.7 Delimitations .....	24
2.8 Research Gap .....	24
2.9 Chapter Summary .....	25
<b>3. RESEARCH METHODOLOGY .....</b>	<b>26</b>
3.1 Introduction .....	26
3.2 Research Philosophy .....	26
3.3 Research Paradigm .....	27
3.4 Research Approach .....	27
3.5 Research Method .....	28
3.6 Research Strategy .....	29
3.7 Theoretical Framework .....	29
3.8 Time Horizon .....	30
3.9 Data Collection and Analysis Method .....	31
3.10 Ethical Considerations .....	31
3.11 Validity and Reliability .....	32
3.12 Chapter Summary .....	33
<b>4. FINDINGS AND ANALYSIS .....</b>	<b>34</b>
4.1 TM Knowledge Patterns .....	34
4.2 Nomenclature .....	37
4.3 Theme of Nature and Wilderness .....	38
4.4 Theme of Mystery and Fear of the Unknown .....	42
4.5 Theme of Superstitions .....	46
4.6 Theme of Death .....	47
4.7 Theme of Community and Shared Responsibility .....	50
4.8 Theme of Loss and Sorrow .....	52
4.9 Theme of Emotions, Sensations and Nostalgia .....	57
4.10 Theme of Natural Transformations and Environmental Changes .....	58
4.11 Theme of Impermanence and Transience .....	60
4.12 Theme of Journey .....	61
4.13 Theme of Human Fragility .....	62
4.14 Theme of Communication and Knowledge .....	64
4.15 Theme of Sex and Marriage .....	64

4.16 Patriarchy versus Matriarchy .....	65
4.17 Significance of Anatomy .....	66
4.18 Interplay Between Characters and the Natural Environment .....	68
<b>5. DISCUSSION .....</b>	<b>70</b>
5.1 Summary Tool: Frequency Distribution of Most Frequent Words .....	70
5.2 Cirrus Tool: Word Cloud Visuals and Theme Insights .....	70
5.3 Collocates Tool: Significant Collocations and Linguistic Patterns .....	71
5.4 Trends Tool: Trends in Word Usage .....	72
5.5 Contexts Tool: Contextual Analysis of Terms .....	72
5.6 Integration of Findings with Research Objectives and Questions .....	72
5.7 Novelty in Findings .....	74
5.8 Significance of Stylometry .....	75
5.9 Relevance of Research Methodology with Findings .....	77
5.10 Contribution of Theoretical Framework in Findings .....	78
<b>6. CONCLUSION .....</b>	<b>80</b>
6.1 Reviewing the Current Procedures .....	80
6.2 Important Results .....	82
6.3 Contribution to the Field of Study .....	83
6.4 Future Research Directions .....	84
6.5 Research Implications .....	85
<b>REFERENCES .....</b>	<b>86</b>



## LIST OF ABBREVIATIONS

2D	Two Dimensional
AI	Artificial Intelligence
CALL	Computer-Assisted Language Learning
CL	Corpus Linguistics
CSSLR	Corpus of <i>Sorrows of Sarasvati: The Lost River</i>
DH	Digital Humanities
EDM	Educational Data Mining
ELT	English Language Teaching
IBM	International Business Machines Corporation
IVC	Indus Valley Civilization
KDD	Knowledge Discovery Theory in Data Mining
LSTM	Long Short-Term Memory
MALL	Mobile Assisted Language Learning
ML	Machine Learning
NLP	Natural Language Processing
OED	Oxford English Dictionary
OMT	Operations Management Theory
PDF	Portable Document Format
PLE	Pakistani Literature in English
RTF	Rich Text Format
TM	Text Mining
TPACK	Technological Pedagogical Content Knowledge Model
VT	Voyant Tools
WE	World Englishes
WWW	World Wide Web
XML	Extensible Mark-up Language

## LIST OF FIGURES

Figure 1 . Main Window of Voyant Tools of CSSLR .....	34
Figure 2 . Cirrus/ word cloud Visual .....	35
Figure 3 . Terms Represented through Trends Tool in the Form of a Graph .....	36
Figure 4 . Corpus Collocates Tool .....	37
Figure 5 . The theme of Nature and Wilderness portrayed through Water and Birds .....	39
Figure 6 . The theme of Nature and Wilderness portrayed through Trees .....	40
Figure 7 . Presence of Elements of Nature and Wilderness in Woods .....	41
Figure 8 . Concordance Graph of Woods with Elements of Nature .....	41
Figure 9 . Relative Concordance of Woods with Village and Peacock .....	42
Figure 10 . The theme of Mystery and Unknown through Water, Bird and Body .....	44
Figure 11 . Water, river, Parushni, and body revealing the unknown mystery .....	45
Figure 12 . Theme of Mystery and Unknown with Masa, Virchan and Parushni .....	45
Figure 13 . Theme of Hope and Superstitions .....	47
Figure 14 . The theme of Death through the Terms Died and Die .....	47
Figure 15 . Collocates of Dead .....	48
Figure 16 . Link of Dead with its Collocates .....	49
Figure 17 . Graph of Collocation of Dead through Trends Tool .....	50
Figure 18 . The theme of Loss and Sorrow through River, Village and Ghaghara .....	52
Figure 19 . Graph of Collocation between Village and River .....	53
Figure 20 . Graph of Collocates of Ghaghara with Water, Village, River and Bank .....	54
Figure 21 . Association of Theme of Loss and Sorrow with the Character of Parushni ...	55
Figure 22 . Collocations of Parushni with Virchan, River, and Samroo .....	56
Figure 23 . Felt portraying the Theme of Emotions, Sensations, and Nostalgia .....	57
Figure 24 . Fields portraying the Theme of Emotions, Sensations, and Nostalgia .....	58
Figure 25 . The theme of Natural Transformations through Ghaghara .....	59
Figure 26 . Collocations of Ghaghara Showing Theme of Environmental Change .....	60
Figure 27 . Trends Tool Graph of Theme of Impermanence and Transience .....	61
Figure 28 . The theme of Journey through Dorga and Leaving .....	61
Figure 29 . Trends Tool Graph of Theme of Journey .....	62
Figure 30 . The theme of Human Fragility through Like .....	63
Figure 31 . Term Kiln shows various Themes .....	63

Figure 32 . The theme of Human Fragility through a Graph of Collocates of Like .....	64
Figure 33 . Theme of Sex revealed through Contexts Tool .....	65
Figure 34 . Theme of Marriage through Parushni and Virchan .....	65
Figure 35 . Instance for Patriarchy versus Matriarchy Debate .....	66
Figure 36 . Significance of Anatomy .....	67
Figure 37 . The interplay between Characters and the Natural Environment.....	68
Figure 38 . Graph of Collocation of Water and River through Trends Tool.....	69
Figure 39 . Significant Trends in Collocation patterns of major characters .....	69

## ACKNOWLEDGEMENTS

I am thankful to Allah, without Whose blessing, I would not have been able to complete this difficult task.

I am very grateful to my supervisor Dr. Arshad Ali for guiding me throughout the research process, without whom this research would not have been possible. I give my sincerest thanks to Dr. Arshad Ali who has consistently challenged my ideas to bring out the best in me. He has always given positive and constructive advice and has been extremely helpful in times of need, I am sincerely grateful for this.

I thank Prof. Dr. Muhammad Safeer Awan, Dean Faculty of Languages and all the teachers, faculty and staff at the MPhil Coordination office for their cooperation in the process. I am also thankful to Prof. Dr. Muhammad Safeer Awan and Dr. Saleem Khan again because of their willingness and permission to use his novel for this study. I hope that I have tried to do my best, but there is much more to explore in the novel.

My previous supervisor Dr. Zafar Ullah (University of Education, Lahore) is also a very motivating figure for me because of his ideas, guidance and proofreading of my thesis. This work is also inspired by the innovative idea presented in his PhD dissertation. He helped me despite his busy schedule.

I express profound gratitude and affection to my parents, especially my mother, whose unwavering support and encouragement have been instrumental in my pursuit of this degree, even when I wanted to leave this degree because of NUML's rad-tapism, academic stagnation and absence of learning. I also want to thank my younger brother (Gul Zeb Shafi), peers and friends, who have supported me in all my years of academic journey. When I regretted being admitted to NUML, they were there to motivate me.

I also extend my deepest thanks to my friends Muhammad Hamzah Masood (University of Management & Technology, Lahore), Iqra Maryam Qandeel Arif (COMSATS University, Islamabad), and Maqsood Ali Darwesh, who guided me throughout the research study despite their busy work schedule. All these friends of mine have worked with Voyant Tools before and are great researchers of Corpus Linguistics, Digital Humanities and English Language Teaching. Their work and cooperation helped in refining my work.

Thank you all.

## **DEDICATION**

This thesis is dedicated to Panj Tan Pak, Iman Hussain Alaih Salam, and the Martyrs of Karbala who sacrificed their lives for the cause of truth.

After that, I want to pay tribute to the civil and military defenders of Pakistan.

This thesis is dedicated to all my teachers, particularly Sir Fakhar-ud-Din Shah, who taught me English at elementary and secondary levels for five (5) years. He is one of the most trustworthy, honest, calm and kind mentors I have ever met.

My mother is the motivation behind all my success, who always supported me in pursuing my dreams, and prayed for my guidance and success. This work is dedicated to her as well who always encouraged me in my studies.

# CHAPTER 1

## INTRODUCTION

The application of technology in the disciplines of humanities has changed the world. Technology is causing a revolution in every aspect of life by aiding humans in various fields (Alsulami, 2016). Many technical advancements have made the previously unthinkable conceivable. In today's world, computers and other technological devices play an important and well-accepted role in language research (Wilks, 2004). Technology assists humans by rapidly extracting precise information patterns from a material or book before actual full reading through text mining (TM). TM is a method of converting unorganized data into an organized form to explore the significant patterns and novel perspectives in the data (IBM, 2023). This technique automatically uncovers knowledge patterns, relationships, links, and information from unstructured data (Hearst, 2003).

With the advancement of technology, language research has relocated to an emerging and sophisticated domain in which technological software aids the research in the field of humanities, known as Digital Humanities (DH). In this way, the computers derive high-quality knowledge patterns from unstructured texts using Natural Language Processing (NLP). The field of DH has revolutionized the way researchers analyse and interpret literary texts, providing innovative approaches to uncover hidden insights and explore the intricacies of literary works (Levenberg, Neilson, & Rheams, 2018). The current study embarks on a thematic investigation of the novel *Sorrows of Sarasvati: The Lost River* using Voyant Tools (VT), a TM approach in the realm of DH. This exploration aims to shed light on the underlying themes of *Sorrows of Sarasvati: The Lost River* within the narrative and gain a deeper understanding of the thematic landscape portrayed in the novel.

*Sorrows of Sarasvati: The Lost River* is a captivating literary work that takes readers on a profound journey through life around the river Sarasvati (mentioned as the Ghaghara River in the novel). This novel delves into the depths of human experiences — exploring the various Proto-Dravidian characteristics that shape the existence of the characters. By employing VT as a TM approach in the current thematic investigation, this study aims to enrich traditional literary analysis by incorporating computational

techniques. This interdisciplinary approach allows readers to delve deeper into the text, unveiling underlying connections and patterns that might not be immediately apparent through conventional close reading alone (Jockers & Underwood, 2015). By integrating digital tools, the study enhances the understanding of the portrayal of the themes in *Sorrows of Sarasvati: The Lost River* and delves into the emotional depth of the narrative.

By employing VT, a powerful TM tool, the study seeks to extract meaningful insights from the textual data of the novel, uncovering the thematic elements and illuminating the nuances of emotional turmoil depicted in the narrative. VT offers researchers an array of analytical features that facilitate the exploration of the text in both quantitative and qualitative ways (Durant, 2004). Through tools such as word frequency analysis, keyword extraction, and collocation analysis, the VT provide valuable insights into the prominence and significance of specific words, collocations, and their associations within the novel. This analytical approach enables uncovering patterns, identifying recurring motifs, and examining the textual strategies employed by the translators to convey the experiences of the characters.

In this research, the researcher explores the key themes from *Sorrows of Sarasvati: The Lost River* using VT. The current study is mixed-method descriptive research that combines quantitative and qualitative approaches. The quantitative data comprised significant terms, their frequencies, word cloud visuals, collocations, and trends of the most frequent words in *Sorrows of Sarasvati: The Lost River*, while the qualitative data incorporated descriptive thematic analysis of the most frequent words, their concordance, and their contexts obtained through VT. So, the theoretical framework combined the postulates of two (2) theories for digitized and productive learning, which address both quantitative and qualitative data. This process is called Operations Management Theory (OMT) and was introduced by Fredrick Winslow Taylor, also known as the ‘Father of scientific management’, in 1911 (Taylor, 2016). OMT combines different theoretical practices with the ultimate motive to promote efficiency in producing results (as cited by Masood, 2022, pp. 27-28).

*Sorrows of Sarasvati: The Lost River* is a translation of a famous Pakistani writer Mustansar Hussain Tarar’s novel *Bahao* by Dr. Muhammad Safeer Awan and Dr. Saleem Khan. *Sorrows of Sarasvati: The Lost River* was published in 2021 by the Pakistan Academy of Letters. It is one of the most complex cultural texts written in Urdu. The novel is a complex tale not only in syntactical and semantic terms but also in its narrative

techniques (Awan & Khan, 2021, p. 9). The manuscript is a post-realist fiction where conceptual complexity is linked with narrative complexity. It revolves around re-imagining one of the most significant historical and archaeological enigmas of the Indus Valley Civilization (IVC). Tarar traced pre-Buddhist and pre-Arian history when the Arians occupied fertile plains of the legendary river Ganghara, also known as Sarasvati in the Arian, the Rigveda, and the Mahabharata scriptures, and enslaved local Dravidians for one thousand years (Awan & Khan, 2021, pp. 10-13).

The current study incorporated a mixed-method approach. The present research combined five (5) tools from twenty-nine (29) VT: Summary tool, Cirrus/ word cloud, Corpus Collocates Tool, Trends Tool, and Contexts Tool. The quantitative findings included statistical data of fifty-nine (59) most occurring terms with their frequencies, a PNG image of Cirrus/ word cloud with sixty-five (65) words, collocations of different terms with their frequencies, and trends of graphs of these frequent terms. The qualitative data comprised the description of the themes extracted from the fifty-nine (59) prominent and other words from the Summary tool, a description of Cirrus/ word cloud visuals, the contexts of the most frequent words obtained through TM, the collocation frequencies of these terms, and graphs of trends of terms. This study falls in the domain of Corpus Linguistics (CL) and Digital Humanities (DH) (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996; Jensen, 2014). This study also interlinks the domain of Pakistani Literature in English (PLE) with World Englishes (WE).

## 1.1 Statement of the Problem

Various concerns paved the way for the current study. First, printed literature texts do not thoroughly perform their crucial function in building and transmitting information in the current digital world (Burdick, Drucker, Lunenefeld, Presner, & Schnapp, 2012). Second, in Pakistan, digital tools like VT are not much utilized for literature, linguistic analysis, and learning methods. Third, *Sorrows of Sarasvati: The Lost River* is one of the latest additions to Pakistani English that has not yet been explored linguistically or thematically. Fourth, the novel has unique terminologies of Proto-Dravidian language, words, phrases, and cultural values that encourage deep linguistic analysis. Fifth, the novel is a compilation of stances and themes deeply woven in archaeology, sociology, culture, anthropology, history, and above all, linguistic features of the past, so these are difficult to analyse with a close reading because of a lack of understanding of students about various domains. On the contrary, VT provide user-friendly details to knowledge



seekers and researchers. Sixth, the current research contributes to Pakistani English by analysing the linguistic features of the novel using digital tools.

*Sorrows of Sarasvati: The Lost River* is a novel equipped with unique linguistic features which are not yet explored. Tarar carried out linguistic research to discover the rhythm of speech as it was more than 4000 years ago. Thereupon, the rhythm is achieved by using Proto-Dravidian words in specific instances, as these words were used across the length and breadth of the IVC in the past (Awan & Khan, 2021, p. 12). The novel has complex words and nomenclature, cultural practices, and Proto-Dravidian dialects. It becomes difficult for contemporary human beings to comprehend such profound words and their associated complex themes. Therefore, it is not easy to understand the lexicon, terms, and nomenclature used by Tarar from the culture that existed thousands of years ago. The current study revealed the linguistic features of these words in their relevant contexts.

Technology is more available to the young generation now than before, and resultantly, they are losing interest in studies. First, teenagers spend their time using these devices because they interact with something new, develop social connections and provide entertainment. Second, digital devices allow learners to learn things more efficiently and unconventionally, and this convenient learning is a neglected part of the life of contemporary students. Third, students lose interest in university courses and books because of less entertainment of course books than digital devices. Fourth, information sources are so massive and dispersed that they are unmanageable without attentive reading. TM increases “digital wisdom” to turn these “static” educational materials into dynamic and ongoing data visualization, so students take an interest in digital learning (Prensky, 2012). This work identifies unified, organized, and visual themes from textual data and better uses digital devices for knowledge seeking.

Pakistani society is separated into several economic strata, which affect the learning habits and choices of students in different academic institutions (Chang, 2014). Lankshear and Knobel (2011) divided students into two groups. One group studied with the aid of traditional note-taking and close reading techniques. The second category of students, on the other hand, preferred to learn using technological devices and resources. The first category belonged to the middle class, whereas the latter belonged to a wealthy class who could readily purchase technological gadgets. It makes a difference in the styles and learning habits of these two groups because online education and digital material

encourage learners towards electronic academia. Moreover, it creates an inferiority complex among students with lower incomes. Thus, VT can help all students without socio-economic discrimination to learn knowledge patterns using the internet because of their free access, user-friendly nature, and affordability.

According to Romero, Ventura, Delgado, and De Bra (2007), printed learning materials like books are more rigid and unorganized than digital learning. The published content is in-adaptable, unstructured, or semi-structured. It takes significant learning time and cannot supply the academic demands of learners. Most students remember the offered knowledge and activities with regular repetition. However, their cognitive powers, comprehension and creativity levels do not increase. Moreover, technology addicts students. In other words, “digital natives” (Prensky, 2001, 2012) of the modern era desire to learn about knowledge structures, data visualizations, and independent computerized TM techniques in the comfort of their own homes (White, 2008, pp. 2-3); so, students should be exposed to digital technologies known as Educational Data Mining (EDM). EDM has recently become a multidisciplinary research, with its roots in CL and DH.

## 1.2 Research Objectives

The research objectives of the present research are specific and measurable and are focused on the outcomes of the study (Jones, 2022). Through TM of Corpus of *Sorrows of Sarasvati: The Lost River* (CSSLR), the current project is designed to achieve the following research objectives:

- i. To reveal the frequency of the most frequent words in *Sorrows of Sarasvati: The Lost River* through the text mining Summary tool
- ii. To create word clouds visuals of *Sorrows of Sarasvati: The Lost River* using the Cirrus tool
- iii. To extract co-occurring words (collocations) in *Sorrows of Sarasvati: The Lost River* through the Corpus Collocation tool
- iv. To identify the trends of the most frequent words in the corpus of *Sorrows of Sarasvati: The Lost River* by the Trends tool
- v. To understand the contexts of different words for the portrayal of themes in the corpus of *Sorrows of Sarasvati: The Lost River* through the Contexts tool

### 1.3 Research Questions

The research questions highlight the primary concerns to be answered in research (Kumar, 2020). So, the following research questions are addressed in this research work:

- i. What is the frequency distribution of the most frequent words in *Sorrows of Sarasvati: The Lost River*, as revealed by the Summary tool?
- ii. How can word cloud visuals of *Sorrows of Sarasvati: The Lost River* be created to provide insights into themes using the Cirrus tool?
- iii. What are the significant collocations extracted from *Sorrows of Sarasvati: The Lost River* through the Corpus Collocates tool?
- iv. How the trends of various words in *Sorrows of Sarasvati: The Lost River* are manifested through the Trends tool graphs?
- v. How do different words in the corpus of *Sorrows of Sarasvati: The Lost River* function in their contexts for the portrayal of different themes?

### 1.4 Significance and Rationale of the Study

The current research is justifiable in many ways. First, the novel is full of complex words and phrases following the pattern of Proto-Dravidian dialects; therefore, these VT greatly benefit learners to explore these words, themes, and their linguistic significance. These words and themes were extracted within their contexts to benefit learners and students further. These tools make a text easy for the reader, which is why a person can easily understand, comprehend, and explain the text no matter the complexity of the themes. These tools help understand the given text by extracting these words as preview themes. Therefore, the vocabulary items and the culture that existed thousands of years ago can be explored via these VT.

Second, these tools also have a contributing educational significance. Students stay uninformed of modern computational learning designs, uniform knowledge patterns, significant characteristics, prominent motifs, significant themes, and central ideas of the content in the documents. In reality, the previewing method empowers the observers to handle brief thoughts and ideas before reading the document through TM of documents. There is an educational need for VT to produce academic ease in the learning process through visual and quantifiable information. Information obtained from VT is practical, time-saving, and energy-saving. Information about the collection of documents and big

data can be obtained in different statistical, visual, and chart forms within a minute. This TM tool, which has twenty-nine (29) more sub-tools, has given a modern path to traditional education (Rockwell & Sinclair, 2016).

Third, searching for keywords, collocation, trends, significant themes, or corpora overview is simple with the help of VT, giving rise to a new wave of digital learning. As humans progress, numerous blogs, books, and other sources have focused on big data and knowledge patterns. Voyant-like tools are now required to extract knowledge patterns from vast data. In addition, as institutions encourage digital books, many contemporary students lose interest in paper textbooks. Digital TM techniques are required to link academics and digital breakthroughs to overcome these issues. It is one of the few studies investigating topics using VT. Some researchers have used digital technologies to mine text, analyse, teach, and research English literature and ESL textbooks (Ullah, 2022). Its intellectual contributions demonstrate its relevance because it previews the writer's themes in visual and statistical form. The significance of these tools cannot be denied because these tools have granted free access to knowledge.

Fourth, the primary stakeholders of the current study are students, instructors, and publishing companies. The current research is motivated by computerized, self-paced, virtual, and interactive learning ideas. Acquiring thematic structures is enjoyable, and a visual understanding of vocabulary in a text corpus encourages students to learn. TM helps find relevance, interact with text, and understand digital information. Instructors and students can better benefit from this corpus and virtual images, so they use this digital text in their classrooms and at home. Teachers may utilize self-made corpora in the classroom at institutions. VT are also available to every student and reader because they are easily accessible, user-friendly, and do not differentiate their users into various social classes. Thus, current research may spawn a new wave of digital learning.

Fifth, the publishing sector would benefit significantly from the findings of this research, bestowing them a social significance. They can produce standard aid materials, exercise books, vocabulary lists, collocations, or lexical categories. Interactive Cirrus/word cloud images help students learn and remember more. These word clouds may be included in popular printed and digital textbooks. Thus, these engaging images may improve digital book sales, and also help in the comprehensive understanding of texts. The current study builds a nexus among literature, linguistics, English Language

Teaching (ELT) and computer science. Most importantly, it defines new textbook study trends using data visualization and interdisciplinary methods (Heydenreich, 2019).

Sixth, many modern Pakistani privatized schools and colleges have implemented CALL (Computer Assisted Language Learning) and MALL (Mobile Assisted Language Learning) in their educational settings using multimedia, tabs, and laptops (Khan & Tufail, 2020). Nevertheless, the majority of students lack digital tools, infrastructure, and training; consequently, they require digital text and tools in and out of classrooms. Therefore, the present work is the practical embodiment of TM of Corpus of *Sorrows of Sarasvati: The Lost River* (CSSLR) using VT (VT). These tools are very effective in the learning process. The learning processes are more profound and more prosperous than the types of learning pupils undergo and are subjected to in schools (Gee, 2012, p. 107). Computer-based education disseminates knowledge to pupils while encouraging individual study (Cristobal & Sebastian, 2013). Remote reading and TM of massive data also uncover a text or corpus's valuable knowledge patterns and linguistic features.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Introduction

An analysis of books, academic papers, or other resources which are necessary to collect relevant and informative data about the chosen subject and theory constitutes the literature review. When doing so, the review gives a brief overview, a critical evaluation, and a description of the literature surveyed concerning the field of study or problem being studied (McCombes, 2023). In the present study, the goal of composing a literature review is to present the reader with a summary of the sources investigated when studying the specific subject and to explain to the reader that the investigation fits within the context of a more comprehensive study area (Jesson, 2011). This literature evaluation included an in-depth analysis and a synopsis of previous academic papers and research that had been conducted about the study. Its purpose was to summarise the level of understanding and awareness of the field of study at the time of writing (Fink, 2019).

The capacity of the literature review to give a framework for the study is one of the reasons that this chapter is considered important. It was helpful to do a literature review because it involved identifying the important ideas, concepts, and current research in the area. It assisted in comprehending the current body of knowledge, locating informational voids, and developing questions for investigation and ideas to assist in defining the context and backdrop of the study issue (Ridley, 2012). In addition, the literature analysis of the current literature revealed several gaps in the understanding. By evaluating the current body of published material, topics and tools were discovered that needed more investigation. This literature highlighted shortcomings and gaps in earlier research and explained why the suggested study is so essential in the contemporary era for TM and digital learning of themes (Rudestam & Newton, 2022).

#### 2.2 Conceptual Literature

##### 2.2.1 Text Mining

Text analytics is a modern phrase for TM. Through statistical analysis, TM extracts high-quality data and information relationships. Some TM techniques include

text categorisation, textual grouping, ontologies, term processing, identified entity identification, taxonomies, idea mining, text synthesis, and text-related modelling. TM is the method which transforms un-organized and unstructured data into organized and structure data. TM discovers novel information and patterns from big data and large amounts of text (Hearst, 1999). Therefore, TM seeks to investigate novel and obscure facts and patterns from substantial textual data. Text analytics, a branch of the DH, engages with every linguistic concern relating to terms, including word clouds, term counts, term combinations, term linkages, frequent and distinct terms, vocabulary density, the most prominent terms, and contexts of words with its relation to other words (Rockwell, 2003).

### **2.2.2 Corpus**

Before the invention of technology, the term “corpus” was referred to as the collection of written data on papers that exhibit similar features. McCarthy and O’Keeffe (2010) state the toil behind building the paper corpus of the Oxford English Dictionary (OED) as “These millions of bits of paper were quite literally categorized in an attempt to organize them into a meaningful body of text” (p. 5). The advanced and modern technical innovation has brought an evolution in language as well. Bonelli (2010) observed, “It was not the linguistic climate but the technological one that stimulated the development of corpora” (p. 15). With the headway of tech-gadgets, the field of language has evolved to a modern era where the paper corpus has been replaced with an electronic corpus with the help of specialized software. Corpus enables the application of computational tools and techniques to humanistic research questions, fostering a more data-driven and evidence-based approach to the study of language and literature (Maci & Sala, 2022).

### **2.2.3 CL and DH**

CL and DH are two closely related fields that utilize computational methods and technology to study language and literature. While CL focuses primarily on the analysis of linguistic data, DH encompasses a broader range of disciplines and explores various aspects of humanistic scholarship using digital tools and techniques (Hoover, Culpeper, & O’Halloran, 2014). However, there is a significant overlap between the two fields, and many researchers in DH incorporate corpus linguistic approaches into their work. In recent years, the integration of CL and DH has been particularly fruitful. Corpus linguistic techniques, such as TM, keyword analysis, and collocation analysis, are

employed to investigate large-scale textual collections in DH projects. These methods allow researchers to identify patterns, uncover hidden relationships, and extract meaningful insights from vast amounts of textual data (Maci & Sala, 2022).

#### **2.2.4 History of DH**

Humanities consist of the social science fields, including art, literature, philosophy, logic, and other related fields. The word “digital” was adopted when technology was used in the discipline of humanities. Empiricism is regarded as a defining characteristic of practical scientific fields. However, as technology advanced, humanities also leapt from conceptual to empirical investigations employing technological advances, reasoning, and scientific tools. The Renaissance (1500–1650) was the time when contemporary human understanding of science first emerged. The dissemination of information at that time was significantly aided by translations from Greek, Latin, Arabic, and French into English. At the same time, humanism grew. Later, the invention of printing helped to standardise communication and spread information about the humanities throughout the world. Thus, several humanities subjects, like linguistics, defined as the scientific examination of language, were frequently alluded to as the fields of science (Ullah, 2022).

Our conceptions of understanding, knowledge, and intellect have changed due to the digital age, digital thinking, and technological usage. As a result of this influence, an array of social sciences disciplines have undergone digitization and been renamed computational social sciences and e-social sciences. This change has led to the collecting of enormous volumes of statistical information in the social sciences, which have been accurately and deftly studied to prove the empirical reliability of newly developed fields (Nyhan, Rockwell, Sinclair, & Ortolja-Baird, 2023). Consequently, academia has gained much reputation with the help of computational linguistics, digital humanities, and data mining (Lazer, 2008). In conclusion, the humanities were transformed into sciences via the use of modern tools for data collection and evaluation.

British colonisation had a significant role in the global dissemination of Renaissance ideas. So, there occurred a change in focus away from humanism and towards many other subjects, leading to the emergence of interdisciplinary research. The post-World War II environment also stimulated the establishment of new linguistics programmes and academic centres at several universities throughout the globe. Conventional patterns of knowledge were given new life by DH. Increased industrial



development under Marxism turned academics into the industry. Above significantly, the World Wide Web (WWW) and Internet access have altered the concept of DH in mobile phones and desktop computers worldwide. Consequently, DH became more deeply ingrained in all academic institutions. With the aid of the internet and Wi-Fi, the idea of digital humanities is revolutionising itself all over the globe. Because of it, DH now has a more significant presence in almost all academic disciplines (Haigh, 2019).

### **2.3 Voyant Tools**

Stefan Sinclair and Geoffrey Rockwell are VT developers, a combination of twenty-nine (29) different tools which can operate in thirteen (13) different languages. This user-friendly tool helps scholars see the most recurring words, their contexts, and the frequencies of an uploaded or pre-uploaded document or text. These sub-tools of VT impressively visualize quantitative, and qualitative data analysis (Hetenyi, Lengyel, & Szilasi, 2019). VT is a web-based, open-source, free software package that provides versatile and advanced text analysis, which helps beginner and advanced humanities scholars (Bradley, 2018; Miller, 2018; Uboldi & Caviglia, 2014; Welsh, 2014). This free open web-based software has been used for quantitative text analysis in numerous peer-reviewed articles (Clouder & King, 2015; Steiner et al., 2014; Williams, Inversini, Buhalis & Ferdinand, 2015; Zahedzadeh, 2017).

VT is a free web-based, open-source text analysis tool that offers an interactive way to extract text (Dickerson & Pack, 2014). It is used to read and analyse large text corpora in any format. The low barrier to entry of VT enables new users to rapidly perform and display text analysis, which is one of the benefits of the software (Jones, 2013). VT is used to analyse extensive data quickly. It is an excellent tool for investigating words and their relationships. It automatically produces graphs and visualizations from uploaded corpora, such as word clouds, which are pictures made of words in which the sizes of the words reflect how often they occur (McGowan, 2021). VT is composed of twenty-nine (29) analytical tools and can operate in thirteen (13) languages (Alhudithi, 2021, pp. 43-44).

The tagline of VT is “see through your text”, and therefore, it offers quantitative validation of patterns you can see in a text, enables the researcher to rapidly identify intriguing words or phrases in a significant corpus, and assists in contextualizing patterns in word usage. An online tool such as Voyant at the beginning of a project may help the

researcher discover intriguing patterns in a text that s/he may further investigate (Platte, 2017). One of Voyant's most significant features is that it enables the researcher to input your text in various ways, depending on the content. It can handle data in a variety of formats, such as plain text, portable document format (PDF), extensible mark-up language (XML), and rich text format (RTF) (Welsh, 2014).

VT works on most browsers without the need for any additional downloads or logins. After loading the corpus, VT offers a wide range of methods to examine texts. There are five primary interfaces of VT: Cirrus, Reader, Trends, Summary, and Contexts. Each pane is interactive, and selections made in one pane may influence another pane's presentation (Hirsch, 2012). Cirrus VT finds focal themes, key motifs, notable characters, and central topics of the transferred literary texts and uploaded documents for digitized visual and statistical word clouds based on textual weight with the aid of the computational programming language, Java, in the form of two dimensional (2D) shapes. The visuals create an enduring and substantial impact on memory, aesthetic sense, and cognition. Therefore, these visuals and statistics are easy to remember and provide essential information about the corpus, mainly terms, frequencies, collocations, and trends (Ullah, Uzair & Mehmood, 2019).

### **2.3.1 Cirrus tool**

The Cirrus tool is an essential instrument in TM and a critical VT (Aggarwal & Zhai, 2012; Rockwell, 2016; Rockwell & Sinclair, 2016). The data obtained from Cirrus is easier to remember as it affects cognition and memory, and the data obtained from the word cloud is long-lasting. Cirrus/ word clouds enhanced language teaching and learning processes (Hayes, 2008). Kolb (1981) introduced four (4) types of learning using word clouds: accommodation, convergence, assimilation, and divergence. They improved the focus, motivation, and level of understanding of the students. Word clouds have thus been proven helpful in the education and learning process (Miley & Read, 2011). At the same time, word clouds carved understanding and admiration. They were excellent for brainstorming, extracting ideas, learning new terminologies and previewing. These Cirrus visuals have a lasting memory effect (Asif, Zaidi, & Yasmeen, 2021).

### **2.3.2 Summary tool**

The summary tool provides an overall summary in brief about the corpus. It gives a quick overview of the uploaded corpus regarding the total number of documents used in

the corpus. This tool tells about total words, unique word forms, multiple occurrences of words, vocabulary density, the average length of sentences, and the most frequent words with their respective frequencies. The summary tool shows the longest and shortest documents in ascending and descending order. An approximation of the average word count per sentence is provided in the same section. The highest and lowest vocabulary density (the ratio of the number of words in the corpus document compared to the number of unique words in the document) in the corpus, along with the line graph, is also presented by this summary tool in ascending and descending order (McGowan, 2021).

The summary tool extracts quantifiable results for discussion. It reveals total words, unique words, vocabulary density, average words per sentence, and most common terms (Eder, Piasecki, & Walkowiak, 2017). Additional properties, such as unique words, total word count, word density, and average sentence length, can be extracted through the Summary tool. The summary tool also effectively finds the unique words, vocabulary density and distinctive words in the corpus. This tool is productive in the preview analysis and the distant reading of a document because it extracts themes after the TM of documents and corpora. The summary tool reveals the most significant terms in the corpus along with their frequency in the corpus. The summary tool is important for revealing the most significant terms which then combine with the Contexts tool to manifest preview themes (Rockwell & Sinclair, 2016).

### **2.3.3 Contexts tool**

Users of VT can investigate the contextual usage of particular words or phrases inside a given text or corpus using the Contexts tool, which is a potent feature. By choosing a target word, readers can view the words around it in the text (context), which gives them information about how the term is used and the larger context in which it appears. Researchers can find patterns, co-occurrences, and correlations that may not be obvious from conventional word frequency analysis alone because of the complete picture of word usage. By enabling users to comprehend the semantic linkages and linguistic subtleties, contexts aid in a deeper examination of the textual data and enhance the discovery of themes, motifs, and narrative structures in DH and CL research. The contexts can be shown from 5 to 50 words, which can be expanded to 500 from 5 by clicking on the expand button (+) in a separate contexts window (Torre, 2019).

### **2.3.4 Trends tool**

The Trends tool in the VT is used to build the graphs of the most used words. These graphs make it possible to see the most prevalent themes, significant changes, and core frequencies throughout the corpus. This graph displays the corpus in ten (10) distinct segments, each containing one of the top five (5) most frequent phrases. These terms can be changed, and the users can see the trend of the term they want to inquire about by searching the specific terms in the search bar. It also shows the trend of collocations. It allows viewers to connect and recognize relationships between the frequency displayed over ten (10) separate parts. These words are displayed using a variety of colour combinations, each of which shows the relative frequency of the term compared to other terms within the relevant section (Neha, 2023).

### **2.3.5 Collocates tool**

The Corpus Collocate Tool displays three columns: term, collocate, and count. Collocates are the terms found around each keyword, Terms are the keywords that need to be collocated, and Count is the frequency of the collocates. The most frequent synonyms for the top ten (10) most frequent terms in the corpus are displayed by default. One can add a keyword specification by typing a query term and clicking enter in the Collocates Tool. Additionally, a slider regulates how much context should be used when looking for collocates. The value specifies how many words should be considered at either end of the keyword, resulting in an overall double-sized phrase window. The default configuration for the context setting places five (5) words next to one another. The slider's value ranges from 1 to 30. By clicking the options icon, one can enter a list of stopwords one wants to avoid (Sampsel, 2018).

## **2.4 Research Literature on VT**

TM, a prominent subfield of data mining, has gained significant attention in recent years due to its potential to extract valuable insights from large volumes of textual data. One of the notable applications of TM is thematic analysis, where researchers employ various computational tools and techniques to uncover patterns, themes, and trends within textual corpora. VT, a web-based text analysis platform, has emerged as a popular choice for researchers seeking to conduct thematic analysis efficiently. This literature review explores the use of VT in TM for thematic analysis, examining its applications, benefits, and contributions to the field of research (Szudarski, 2023).

TM, also known as text analytics, involves the automated analysis of large textual datasets to discover meaningful patterns, trends, and knowledge. It encompasses a range of techniques, including information retrieval, machine learning (ML), and statistical analysis, aimed at extracting valuable insights from unstructured text data. Thematic analysis, on the other hand, is a qualitative research method used to identify, analyze, and report patterns (themes) within textual data. It is widely employed in various fields, including social sciences, humanities, and business, to gain a deeper understanding of the content and context of textual sources.

VT is an open-source, web-based text analysis platform that provides a user-friendly interface for exploring and analyzing textual data. Developed by scholars and researchers, VT offers a suite of TM functionalities, making it accessible to both novices and experienced researchers. The platform facilitates tasks such as visualization, summary, concordance analysis, and trends, enabling users to conduct in-depth analyses of textual corpora. VT is particularly valuable for thematic analysis as it allows researchers to uncover underlying themes and patterns within textual data efficiently.

The following part of the study examined some past studies on VT which are relevant and applicable to the current study. Different studies have applied VT in their research for thematic analysis. Archer and Bousfield (2010); Archer, Culpeper, and Rayson (2009); Craig and Kinney (2009); Craig (2010, 2011); Culpeper (2002, 2009); Hendrigan (2019); Hope and Witmore (2004, 2010); Kairaitytė-Užupė, Ramanauskaitė and Rudžionis (2023); Kiselev (2021); Lessa (2020); Miguel (2022); Mireille and Culpeper (2011); Petersen (2010); Schreibman, Siemens, and Unsworth (2008); Scott and Tribble (2006); and Sinclair (1996, 2004, 2007) used VT and corpus methods to the linguistic investigations of different literary and scientific work.

Jannidis applied VT to 350 Canonical German books. This investigation was restricted because of its inability to indicate each theme for its readers about every book (Burrows, 2002, p. 267). Using VT, Sinclair and Rockwell (2015b) conducted a similar analysis of the Jane Austen corpus. It dealt with an incredible amount of textual information. In the study, comparatively interesting, informative, and unique patterns were inferred in statistics and data visualization—word clouds from these fictions. Similarly, Daines, Nimer, and Lee (2018) conducted a study with the help of VT to explore the American archivist. The suggested corpus analysis and its supporting tools

were a successful and novel approach to studying the expert archive literature, which is considered advantageous and beneficial for learning.

Ullah, Uzair and Mehmood (2019) applied VT to the Nobel prize-winning novel *Never Let Me Go* to extract the key motifs. In this research, the main goal was to create an interactive word cloud based on statistical data to give an overall picture of the whole novel through previewing techniques with the help of VT: Cirrus, Context and Statistical Tool under the KDD Model. In 2016, Ramsby studied the non-fiction writings of two prominent authors: Zora Neale Hurston and Richard Wright. TM techniques have been used to analyse several short stories by these authors selected for analysis. TM software, VT, assisted in recognizing linguistic characteristics, noteworthy trends, and other patterns. These studies provided a platform for understanding the usage of VT.

Due to the sheer usefulness of VT, several studies were carried out to investigate valuable knowledge patterns. Utilizing VT, Yeates (2013) created word clouds from a large dataset of 1500 apocalyptic books using big data. VT were also used in periodicals to investigate their central themes (Grier, 2014). In addition, A Ph.D. student, Kemman (2016), from Luxembourg University, applied VT to analyse thousands of WikiLeaks emails and found the answers to who, what, when, and where. He conducted the content analysis and found the incorporated visual representations, themes and patterns in those emails. These studies were a significant contribution to the domain of CL, DH and data mining, but were limited to a single VT only.

ReCloud, a semantic word cloud, is the integrated context in the cloud world. Consumer happiness was assessed utilizing the semantic cloud in business intelligence. However, the Context tool revealed more interactive context of essential words than ReCloud (Wang, Zhao, Guo, North, & Ramakrishnan, 2020). Reviews from Yelp and Amazon were gathered using various techniques. The context tool of VT is more effective and advanced for finding contexts (Ullah, 2022). Many different word clouds have also been constructed by computer technology, such as Cirrus, which was built using the Wordstorms Toolkit to extract data patterns from text. A sound word cloud would not overlook an essential topic or character while filtering out stop words (Castellà & Sutton, 2014). The Voyant tool has been equipped with the two characteristics stated above; thus, its outcomes are standard and polished.

VT could also be used for dynamic analysis beyond theme extraction. Some studies extended word clouds on emotion mining. Shakespearean dramas *As You like it* and *Hamlet*, Shelley's *Frankenstein* and a few more tales were explored to uncover the patterns of emotional knowledge (Mohammad, 2012). Cirrus picked up emotional topics to uncover psychoanalytical character conditions since dramatic genres revealed mental problems and emotional states. Most literary works incorporate emotions, so the Cirrus Voyant Tool is needed to obtain accurate and reliable results (Ullah, Uzair & Mehmood, 2019). Some other researchers focussed on using content analysis (Krippendorff, 2003); however, it revealed insufficient fluidity. The era has moved to accurate, mathematical results and quantification, more than a qualitative approach. Therefore, the quantification is included in the current study which leads to the thematic analysis.

There were several literary controversies over the authorship of literary and non-literary works of some writers. The stylistic analysis of texts via VT might be used to settle these disputes, as several previous research done (Stamatatos, Fakotakis, & Kokkinakis, 2000; van Halteren et al., 2005). The text and thematic analysis of an author finds the key writing characteristics of a particular author. Moreover, the writings and styles of Dutch students in 1126 documents called CLiPS, which contained 305,000 token words, discovered personality, feelings, author, age, gender, and genre (Verhoeven & Daelemans, 2014). Average sentence length, vocabulary density, and the most used words were not taken into account in this study. Hence, the summary tool filled this gap.

Using VT, Li, Ji, and Xu (2017) produced a literary corpus of several works by Mo Yan and Zhang Wei to distinguish between their stylometric measures. Statistics revealed that their writing style varied in sentence structure, dictionary length, and social system description. The language of Zhang Wei changed and often varied, whereas the language of Mo Yan stayed constant throughout his works. In the same way, the literary work of Ernest Hemingway was reviewed to identify prevalent theories and stylometric characteristics in the characterization of female characters in Sweden. Thus, names, tokens, and adjectives were listed to define stylistic structures with the help of the Summary tool (Nilsson, Sundberg, & Höglund, 2018).

The VT were based on textual weight and statistical data. Consequently, the most occurring words showed a prominent and noticeable spot. VT portrayed significant results because of its rectangular shape with its ability to combine word clouds and their obtained statistical information with the help of its occurrence numbers. So, the Voyant Cirrus tool

proved superior to other digital tools (Ullah, Uzair & Mehmood, 2019). Thus, applying VT to most works proved beneficial for exploring new knowledge domains. It produced fast memory-oriented results with the help of Summary tool statistical data and patterned visual learning aided by Cirrus/ word clouds. These applications of VT have gained importance in education and teaching purposes (Sinclair & Rockwell, 2015a, 2015b, 2015c).

Using VT, Miller (2018) researched at Middle Tennessee State University, evaluating one of the most current projects, *Trials and Triumphs*. The author intended to serve as helpful advice for librarians and other researchers interested in creating DH initiatives. The findings of this test were thoroughly discussed. The tools the researcher chose included Bubblelines, Cirrus, Corpus Collocates, Contexts, Corpus Terms, Knots, Documents, Reader, Scatterplot, StreamGraph, Summary, Terms Radio, and Trend VT. Using VT to analyse the content helped define keywords, trends, frequency and relativity. It finally resulted in the identification of contemporary study fields. VT visualisations helped in this area since they helped to steer and guide research in ways that were not achievable using conventional research methodologies.

Tokarz (2017) used Voyant and NVivo 12 to analyse key words and phrases in the article titles and abstracts to determine the study areas covered by faculty publications in peer-reviewed journals. Science web was used to get publication data from the University of Nevada's Department of Criminal Justice. For this particular investigation, 176 publications in total were chosen. The researcher obtained better results by omitting stop words and other keywords and phrases from analysis, such as phrase analysis. Some phrases were retrieved from the data using VT based on their frequency of occurrence. However, the group analysis of NVivo 12 synonyms could not provide any fresh, widely used terms that may be used to highlight possible study interest areas.

*Romance of Flamenca* and its English translation were examined using the Toolbox of Interactive TM Suite. The study focused on using word clouds to analyse important ideas, core motifs, key topics, and notable personalities. This computational study comprehensively depicted the information connected to close, inaccessible, and distant reading using DH and CL (Scrivner & Davis, 2017). Given that the previous study was about novels and the current study is about poetry, there were no apparent similarities between the two genres. There was much overlap with the word cloud, but a separate method was used for the study. The Voyant technology under examination combined DH



with corpus to investigate novel and unique word cloud designs to aid in knowledge discovery. Other researchers also conducted similar studies using Cirrus/ word cloud (Godwin-Jones, 2006; Castella & Sutton, 2014).

VT can be employed to quantify a large amount of text or big data (Stamatatos, Fakotakis, & Kokkinakis, 1999). All the works of Roberto Busa may be analysed in a few minutes using VT (Argamon, Šarić, & Stein, 2003; Zhang, Damerau, & Johnson, 2001). There are many tools readily accessible for corpus analysis. Few people have used VT because it is a relatively new technology. There are some instances in academic publications where VT are utilised to analyse language corpora other than English. Among such papers is Lorie A. Vanchena's "Reading German Culture, 1789-1918, " which examines the contextual usage of the words "Love and Death" in German literature (Vanchena, 2012), and Jeffrey Drouin's work "Close and distant reading modernism: Network analysis, TM and teaching the little review" (Drouin, 2014). Lessa (2020) and Torre (2019) conducted similar studies on Spanish language corpora applying VT.

Altameemi and Altamimi (2023) introduce an innovative approach to thematic analysis through the utilization of Long Short-Term Memory (LSTM) networks. This method offers a novel perspective on understanding the themes and topics within a textual corpus by employing a classification process. By harnessing the power of LSTM, a type of recurrent neural network known for its ability to capture sequential patterns in data, the authors present a promising avenue for automated thematic analysis. This research marks a significant contribution to the field of text analysis, particularly in the context of large-scale textual datasets, and demonstrates the continuous evolution and adaptation of methodologies in the digital age.

Ullah (2022) further extends the theoretical landscape by applying a Triangulation approach that employs the operational model, combining Knowledge Discovery Theory in Data Mining (KDD) with the Hermeneutica Theory. This synthesis of knowledge discovery and hermeneutics reveals common features and establishes a fundamental relationship between the two, underlining their complementary roles in understanding textual data. By integrating KDD and the Hermeneutica Theory into the current TM study, the researcher aimed to explore knowledge patterns (Cristobal & Sebastian, 2013) within the textual corpus. KDD, defined as the extraction of implicit, previously unknown, and potentially valuable information from data, is a pivotal component of this approach. It serves as a bridge between unstructured data and structured, valuable information,

aligning with the broader goals of digital tools and theoretical frameworks in DH. This synthesis not only enhances the robustness of the research but also demonstrates the transformative potential of theoretical frameworks in the realm of digital TM.

VT has emerged as a valuable resource for researchers conducting thematic analysis through TM. Its accessibility, efficiency, and flexibility make it a preferred choice for exploring textual corpora and uncovering hidden themes. As the field of TM continues to evolve, VT remains a powerful ally for researchers seeking to extract meaningful insights from textual data, contributing to advancements in knowledge across various disciplines. Its applications in humanities, social sciences, business, and beyond demonstrate its versatility and potential in enhancing the rigour and depth of thematic analysis in this research.

## **2.5 Literature on OMT**

Fredrick Winslow Taylor, commonly known as the ‘Father of Scientific Management,’ made significant contributions to the field of management by demonstrating that a scientific approach could be applied to enhance organizational efficiency. Taylor’s practical theories were deeply rooted in the operational management of research and performance at the ground level of research. It believes in combining the postulates from more than one theory for scientific management and operationalization of analysis (Masood, 2022). He placed a strong emphasis on optimizing and simplifying job tasks as a means to boost productivity. Taylor’s foundational idea was that by carefully studying and refining work processes, organizations could achieve higher levels of efficiency and output (Byjus, 2019).

One of Taylor’s key insights was the importance of cooperation between theories in achieving research-driven and productive outcomes. Taylor’s scientific management theory underscored the significance of evidence-based methods in management. By relying on empirical evidence and systematic analysis, researches could prioritize efficiency and reliability in their analysis (Jameslopresti, 2022). This approach marked a departure from traditional theoretical practices, which often lacked a scientific basis. When Taylor’s principles are properly implemented, they have been shown to improve productivity and efficiency within organizations. Taylor’s legacy has endured, influencing modern management practices and providing a foundation for evidence-based management methodologies (Tools, 2023).

Masood (2022) conducted research using VT on Shakespearean Corpus. The research draws on an operational model that integrates concepts from multiple frameworks for digitized learning. Specifically, it leverages the Knowledge Discovery Theory in Data Mining (KDD) which was proposed by Rakesh Aggrawal. This approach aims to unearth valuable and novel patterns of knowledge and learning designs through the evaluation and interpretation of visual data. Additionally, the study incorporates the Interactive Word Cloud Model (IWM), as proposed by Carmel McNaught and Paul Lam. This model posits that the insights derived from the statistical word cloud visuals are as reliable and valid as those obtained through manual examination. These theoretical foundations serve as the framework for analyzing the digital text in the research.

The analysis employs an operational learning model that synthesizes principles from Post Positivism and Epistemology (John, 2017). Within this framework, the study adopts a descriptive research strategy, harnessing the power of artificial intelligence (AI), deep learning, and ML tools. A distinctive feature of this research is its utilization of a mixed-methods approach, seamlessly blending qualitative and quantitative methods. This interdisciplinary approach extends to various disciplines, including AI, deep learning, ML, and hermeneutics (Fayyad et al., 1996). An integral component of the study is the application of the Technological Pedagogical Content Knowledge (TPACK) Model, proposed (Koehler & Mishra, 2009). This model underscores the innovative integration of technology, instructional methods, content, and pedagogy within the learning process (Flick, 2014). It serves as a guiding framework to explore the dynamic relationship between technology and education, shaping the methodology of the research. This model was applied to create innovation in the learning process and build a nexus among the different domains of study (Cabena et al., 1998a, 1998b). This model was an operational research model applied in these researches for the productivity of the results and analysis.

This literature review highlights the theoretical underpinnings that inform the research. The research adopts a mixed-methods approach, drawing from diverse disciplines, to explore the intricate intersection of digitized learning, technology, and knowledge discovery (Ullah, 2022). These theoretical foundations collectively contribute to the research's innovative and multidimensional analysis of digital text, furthering our understanding of how technology and learning intersect in the digital age. The operational model is useful in building a nexus among different domains of study. An operational model manages the research by combining postulates of two distinct theories and

combining them to produce effective results. In this model, one theory leads the data towards another theory (Masood, 2022).

## 2.6 Limitations

In any research endeavour, it is essential to acknowledge the limitations encountered during the study, as these constraints can influence the interpretation and generalizability of the findings. Since the translators employed various synonyms throughout the novel, reusing the exact words or inferring themes from statistical value can be unsuccessful or misleading. Occasionally, a writer also alludes to a concept without explicitly using the phrase that best captures it. VT are technologically limited because they cannot receive and publish InPage text. Only a few prior studies were determined to be pertinent to VT. Multiple individuals had referenced VT in blogs or lecture notes. The VT sometimes skip the essential themes presented through synonymous words, their inflexions and derivatives. Some terms, which are essential for the research, are sometimes skipped by the VT based on their non-concrete nature in the corpus. These limitations are the hindrances faced in the current analysis, but still, these tools were advanced, user-friendly and convenient.

Several other limitations have been identified. First, the research heavily relies on the availability and quality of the source text. Any inaccuracies or ambiguities in the text can impact the accuracy of the analysis. Second, the analysis is focused on five (5) TM tools and methodologies. Other analytical approaches or tools that could provide additional insights may not have been explored. Third, this research is limited by the language and cultural context of the text. Interpretations and themes may vary depending on the reader's familiarity with the cultural nuances and linguistic subtleties. Fourth, findings from this study are specific to *Sorrows of Sarasvati: The Lost River* and may not necessarily apply to other texts or literary genres.

Some other limitations of eminent significance are that the VT are only able to distinguish and reveal the preview themes from the corpus through TM. It means that *Sorrows of Sarasvati: The Lost River* might have some other descriptive themes which can only be revealed after a thorough reading of the novel. It also indicates that the themes which are extracted are based on the quantification and frequencies, their visuals and links, their collocations and connections, their trends and graphs, and their relative contexts. The themes are extracted based on the five (5) aforementioned tools from VT.

These tools extracted the words using these VT, and then the thematic analysis was carried out based on distant reading. VT provide insight into the corpus and provide insight into the hidden and novel patterns in the unstructured literary data.

## 2.7 Delimitations

The empirical results herein could be given with caution of certain delimitations. The study is limited to analysing *Sorrows of Sarasvati: The Lost River*, translated by Dr. Muhammad Safeer Awan and Dr. Saleem Khan. The analysis is carried out with the help of statistical data obtained from only five (5) VT: Summary tool, Cirrus/ word cloud, Corpus Collocates tool, Trends tool, and Contexts tool. This study is delimited for TM and thematic analysis of only the translated version of the novel *Sorrows of Sarasvati: The Lost River*. The inquiry only considers the operational theoretical framework: A quantitative study was conducted under digital TM (Ignatow, 2015, p. 22), and the qualitative study incorporated thematic analysis (Gibbs, 2012, p. 2).

When defining the parameters of this study, it is essential to make it clear that the primary emphasis is on the analysis of themes using computational methods. The research aims to clarify thematic components directly obtained from textual analysis, excluding suggestive themes from its scope. However, it is important to recognize the inherent limits of the instruments used, especially when it comes to capturing the qualitative subtleties found in the text. Although computational techniques such as VT provide insightful information on word frequencies, collocations, and thematic patterns, they are unable to adequately convey the depth of qualitative data present in literary works. As a result, context tools and close reading techniques are employed in the research to cover the qualitative part of data analysis.

## 2.8 Research Gap

The research gap was present in the previous studies, particularly about the VT and the novel *Sorrows of Sarasvati: The Lost River*. The research gap lies in the limited application of TM tools like VT in the analysis of themes, specifically through TM, within literary texts. While DH research had significant advancements in the analysis of literary works using computational methods, there remains a scarcity of studies that specifically explore the thematic elements in literature. Existing research has primarily focused on more general themes or topics, neglecting the intricate portrayal of terms, themes and their impact on the narrative. This study aims to bridge this research gap by

employing VT as a TM approach to uncover the underlying themes within the novel *Sorrows of Sarasvati: The Lost River*. Through this thematic investigation, the study seeks to demonstrate the potential of DH tools in exploring thematic landscapes within literary texts and contribute to a deeper understanding of the emotional dimensions of storytelling within the novel.

In addition, *Sorrows of Sarasvati: The Lost River* was not analysed for its themes or any other literary analysis, so this study is a significant contribution to the field of PLE, DH, CL, and translation studies. *Sorrows of Sarasvati: The Lost River* were also not examined with the help of digital tools, particularly with the help of text analytics. In the current analysis, the operational theory was not applied to the *Sorrows of Sarasvati: The Lost River* to study it extensively through TM. Statistical analysis is seldom conducted to get insights into particular themes, raise new inquiries, and even reconsider academic assertions, which is an innovation in the current analysis. The VT is least applied to literary corpora in developing countries, including Pakistan, and the link between humanities and digital technology is missing in the learning process in previous studies, and this study helps to fill these gaps.

## **2.9 Chapter Summary**

Existing information must be cited while building a unique understanding since it builds on earlier research findings and moves from the familiar to the unfamiliar. The current part provides the historical development of specific fields, including DH, CL and PLE, which is directly related to ongoing research. Additionally, the data analysis section utilized this segment to do an additional check and explain if the results of the present study agreed with those of prior investigations. It strengthened the present research and demonstrated its relation to previously published research. This chapter included significant advancements and studies regarding the five (5) VT, data extraction, TM, knowledge patterns, graphics, thematic analysis, trends, collocations and contexts. Therefore, it spanned every aspect of the scientific and educational approach, from the general concept of DH to the modern Voyant TM techniques.

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1 Introduction

The research methodology provides a direction for the research. This chapter is the methodical and logical process of gathering information to solve specific research issues using various methods and approaches, and a complete research process (Kumar, 2020). The present research methodology aims to produce fresh and thoughtful perspectives on current scientific paradigms in general. For the current investigation, the research methodology is designed according to the research onion (Saunders, Lewis, & Thornhill, 2012). It includes research philosophy, research paradigm, research approach, research method, research strategy, time horizon, theoretical framework, data collection method, data analysis method, ethical considerations and validity and reliability. This chapter is the intellectual framework and the basis around which the focus of the current investigation is built and provided the process for the present research.

#### 3.2 Research Philosophy

The current research study adopted post-positivism (post-empiricism) philosophy. Post-positivism is a research philosophy that acknowledges the subjectivity of researchers while emphasizing the importance of empirical evidence and objectivity in the research process. It accepts that interpretations of people are influenced by their perspectives and biases but also strives to maintain objectivity through rigorous data analysis and validation (Bhaskar, 2011). In the context of the thematic investigation of the novel *Sorrows of Sarasvati: The Lost River* using VT in DH, the post-positivist research philosophy guides the research. Post-positivism highlights the importance of empirical evidence in research. In the thematic investigation of the novel, this means that the conclusions drawn about the thematic elements related to sorrows were based on concrete evidence derived from the textual data (Blaikie, 2010). The use of VT in TM allows for the quantitative analysis of the text, providing empirical evidence to support the identification of themes and patterns.

Post-positivism does not adhere to relativism and often upholds objective truth. In ontology, post-positivists think that reality exists yet that it is only capable of being partially understood (Brown, 2006). So, Post-positivism recognizes that researchers' subjectivity can influence their interpretations. However, it advocates for objectivity in data analysis and interpretation. By employing Voyant Tools, the research process is guided by quantitative analysis, reducing the potential impact of individual biases and interpretations. The tool's data-driven approach ensures a level of objectivity in the analysis of word frequencies, keyword extraction, and collocation patterns. Moreover, post-positivism values the generalizability of research findings. In the context of the thematic investigation, this research drew conclusions which were applicable beyond the specific novel being studied. By using VT, which can process large volumes of text data, the study can potentially uncover patterns and themes that have broader relevance in DH (Cohen, Manion, Morrison, & Morrison, 2007).

### **3.3 Research Paradigm**

Burrell and Morgan (2016) said that a paradigm is an arrangement of fundamental and widely accepted presumptions that serve as the foundation for an entire population's operational frame of reference, manner of theorizing, and research methods. The research paradigm of the present study is interpretative. The primary subject of a study conducted here is how people strive to become acquainted with the contexts and environment around them. This paradigm believes that there can be multiple subjectivities which can be uncovered with research methods (Kelemen & Rumens, 2008). The primary objective of the interpretative paradigm was to comprehend the underlying significance of themes within CSSLR using VT. It emphasizes the subjective analysis of the data instead of logical conclusions, and the current study deals with the subjective analysis of themes.

### **3.4 Research Approach**

The current study adopted an abductive research approach (premises) since it uses VT to conduct digital TM and thematic analysis to produce intriguing knowledge patterns. This research approach is utilized because it works as a top-down as well as a bottom-up approach. It not only combines but also addresses the shortcomings of both inductive and deductive approaches (Creswell & Clark, 2017, p. 23). Throughout the study process, the impacts of hypothetical inquiries, theories and research tools were tested in this method and incorporated into existing theories. Five (5) research questions were answered using



five (5) tools in the current study: First, the Summary tool generated corpus summary; second, the Cirrus/ word cloud visualized themes; third, Corpus Collocates Tool disclosed the significant collocations; fourth, Trends tool portrayed the trends of most frequent words; and lastly, Contexts Tool revealed the contexts of various terms in the novel. Various unknown knowledge patterns were revealed through digital TM using VT.

### **3.5 Research Method**

The data for this study is collected through both qualitative and quantitative methods. So, the current research framework addresses the mixed-method approach or “combination of qualitative and quantitative methodologies” (Durant, 2004, p. 10). English, ML, applied linguistics, NLP, EDM, AI, text analytics, and statistics were all included in this research project that combined linguistics and computer sciences. Thus, it eventually changed into a multidisciplinary investigation (Walliman & Walliman, 2011). The research design aligns with the theoretical framework because each postulate of operational theory explains one aspect of the mixed method: quantitative and qualitative.

The mixed method approach was utilized in this research because of various reasons. Compared to a single technique or research method, their triangulation allowed a more nuanced, complete, genuine, and deep understanding of research problems (Creswell & Clark, 2017). It is because the quantitative content included the statistical frequency of documents, words, and vocabulary density, whereas the qualitative content, such as bidirectional context, themes taken from word clouds, and visual method, were analysed. The present study uses mixed methods rationally since data analysis uses digital tools (Surbhi, 2016). As cryptographic machines, computers generate “numbers, colour, characters, and so forth” (Ullah, Uzair, & Mahmood, 2019). Thus, the mixed-method approach is rationalized because one method is insufficient for the current study design.

It was recognized that thematic or qualitative data is an essential component of data—additionally, computers “substitute quantification for intuition” (Masood, 2022). As a result, mixed approaches were the best suitable for data collection and analysis. It made results valid and reliable. Firstly, pragmatism favoured the hybrid mixed-method approach. Secondly, the Voyant TM tool produced both qualitative and quantitative data, and a mixed-method approach is required for the data collection and analysis. Thirdly, the objectives of the present study required a mixed-method analysis approach. Fourthly, data mining tools aided hybrid learning (Adriaans & Zantinge, 2009).

### 3.6 Research Strategy

The research strategy for this study is descriptive thematic analysis. As the thematic investigation aims to uncover underlying themes within the narrative, the descriptive research strategy allows for an open-ended and flexible analysis of the themes obtained through VT (Gregory, Geiger, & Salisbury, 2022). It has enabled the researcher to delve into the textual data using VT without being bound by preconceived hypotheses (Mountain & Marshall, 2019). The descriptive approach is thus effective in this thematic investigation because it allows the researcher to identify patterns, themes, and relationships in the text, paving the way for deeper insights into the emotional landscape portrayed in *Sorrows of Sarasvati: The Lost River*. Although this study has an objective ontological basis, knowledge is presented via personal understanding (John, 2017).

Thematic analysis is a versatile approach that may be used in various studies. Various ways of thematic analysis exist, but the most popular type follows a six (6) step process: Familiarisation, coding, generating themes, evaluating themes, defining, and labelling themes, and writing up the study results (Caulfield, 2019). Virginia Braun and Victoria Clarke first devised this procedure for psychological research. In the current study, the semantic thematic approach was followed to analyse data for its explicit content rather than structure (Braun & Clarke, 2021). Thus, it is an approach within the semantic thematic analysis that deals with the deep descriptive analysis of themes. In the present study, the descriptive analysis of themes is completed through TM using VT.

This subjective experience can be humanistic or produced via digital tools. In the current study, the thematic analysis is completed using TM. It is because when doing a thematic analysis of documents, both theoretical or analytical examination, as well as a textual description, are important to be taken under consideration. During this procedure, one should read the material extensively to guarantee that one recognizes all the essential concepts in the text, even the most basic. In the case of the current study, this extensive analysis is completed by the TM technique, which was carried out through VT (Gao & Wallace, 2017). Thus, the results are generated in the form of themes, which were obtained through the description of contexts of various frequent terms.

### 3.7 Theoretical Framework

A framework provides the support and guiding structure for research (Kumar, 2020). The theoretical framework applied in this study was an operational model that

combined postulates of two (2) theoretical frameworks for digitized learning (as cited by Masood, 2022, pp. 27-28). This process is called Operations Management Theory (OMT) (Taylor, 2016). This model was given by Fredrick Winslow Taylor, also known as the ‘Father of scientific management’. Thereupon postulates of two (2) theories were combined in this research under OMT: Digital TM Theory and Thematic Analysis Theory. First, quantitative analysis follows digital TM theory {an extension of Digital Hermeneutics Theory by Rockwell and Sinclair (2016, p. 116)}, stating, “digital TM resources can be used not only to amplify human interpretation through sheer computing power but to supplement human interpretation through applications of sociologically and linguistically sophisticated algorithms” (Ignatow, 2015, p. 116). Second, the qualitative study incorporated thematic analysis theory which states that “when doing qualitative research, thematic analysis is used to recognize or capture text or pictures connected by a similar idea or concept. It allows a researcher to classify the content and, as a result, create a ‘framework of thematic ideas’ about it” (Gibbs, 2012, p. 2).

As the operational model is used as the theoretical framework in the current study, postulates of digital TM and thematic analysis were applied to find unfamiliar knowledge patterns and learning designs through the evaluation and interpretation of visual data. It is the extraction of implicit, previously unknown, and potentially useful information from data called a semantic thematic approach. Thereupon, raw data is transformed into meaningful and valuable knowledge patterns; for example, the Cirrus tool shows themes; the Summary tool exposes details about the corpus, unique words, and stylometric features; the Corpus Collocates Tool discloses the significant collocations; the Trends Tool portrays the trends of most frequent words; and the Contexts Tool reveals the contexts and concordance in the novel.

### **3.8 Time Horizon**

The time horizon for this study is cross-sectional. A cross-sectional study gathers information from a sample at one particular period in time. It is because when the data is collected at a single point in time, it makes research more efficient and reliable. In addition, this time horizon adopted an epistemological viewpoint, in which nothing is regarded as an absolute reality because reality is a relative term (Driessen, 2021). In this analysis, the themes were collected once through TM of the corpus using VT. The data is collected from *Sorrows of Sarasvati: The Lost River* single time, instead of repeated inquiries of data gathering. Thus, the time horizon for the current study is cross-sectional.

### 3.9 Data Collection and Analysis Method

This part of the study elaborates on how themes were collected and applied. First, the corpus of *Sorrows of Sarasvati: The Lost River* was uploaded in VT. Second, a new window of five (5) panels with five (5) tools was opened. Third, the fifty-nine (59) preview themes with their frequencies were obtained from the panel of the Summary tool. Fourth, the Cirrus/word cloud PNG picture was imported to the Word file as sixty-five (65) visual themes. Fifth, the statistical content with relative frequencies of words and Cirrus visuals were exported to a Word file. Sixth, the collocations, trends, and contexts were exported to the Word document, relevant to the thematic analysis. Seventh, themes were obtained from the most frequent terms, their contexts, concordance and trends. Lastly, these themes were analysed quantitatively and qualitatively using terms, contexts, concordance and trends.

### 3.10 Ethical Considerations

In conducting this thematic investigation, ethical considerations were of utmost importance. It was ensured that the proper citation and acknowledgement of the original translators' work was given when using the textual data from the novel and throughout the research process. In addition, the research was conducted with respect to the characters, contexts, and themes of the novel, so the researcher approached the analysis with sensitivity while exploring themes. Appropriate ethical safeguards were put in place to protect the confidentiality and privacy of the data involved. The knowledge patterns generated through CSSLR are used for research purposes only and are not utilized elsewhere. This investigation does not focus on any special considerations or permissions because it does not include human subjects.

The VT is open source and freely available to the public. As a result, permission is not necessary for using VT. Second, VT is also accessible to the public and downloadable. Open access means anybody may use it without prior permission or payment. Third, the themes generated in this study can be used for publication or verification. Fourth, before conducting this study, permission was taken from the primary translator of the novel Dr. Muhammad Safeer Awan to use the novel *Sorrows of Sarasvati: The Lost River* for research purposes, and the novel would not be used for any financial purposes. Lastly, the readers of this research can use the knowledge patterns and themes generated in the current study without prior permission for academic purposes.

Sensitivity, confidentiality, and privacy of data are critical considerations in any research project, including the present research involving TM and analysis of *Sorrows of Sarasvati: The Lost River*. Sensitivity, confidentiality, and privacy considerations in the present research are crucial to ensure ethical research practices. Researchers must be diligent in safeguarding the data they work with, especially when analyzing textual content that may contain sensitive or personal information. Adherence to ethical guidelines and practices is essential to protect the rights and privacy of individuals and entities associated with the data.

While the primary focus of the research is on TM and analysis, it is important to be aware that the text might include elements that could be sensitive in nature. This could include details about characters, settings, or cultural references that are sensitive to certain groups or communities. Depending on the source of the text, as the novel is a translated published work or provided by a third party, the researcher has ensured that the data is stored securely and is accessible only to authorized personnel involved in the research. The researcher has an ethical obligation to maintain the confidentiality of the novel. Any data collected or accessed for research purposes would be handled with care and not disclosed to unauthorized individuals or entities. The text being analyzed involves user-generated content or data obtained from translators, the researcher has obtained proper informed consent from those individuals, ensuring that their trust is respected, and the soft-copy of the translated work would only be used for academic and research purposes.

### **3.11 Validity and Reliability**

As the methodology is a reason for selecting one or more correct research techniques, which should be validated following the reliability and validity standards; therefore, several measures were taken to ensure the validity of the research findings. Firstly, textual data was carefully selected and analysed to establish internal validity. It was ensured that the study accurately represents the content and themes of the novel obtained after TM through VT. Internal validity is also achieved through careful selection of content of the novel obtained through VT and qualitative thematic analysis. Secondly, external validity was addressed by documenting the research methodology, data sources, and analytical techniques. Thus, the study ensured internal and external validity in the present research by applying appropriate methods for data collection and analysis.

The reliability of the research was also ensured by maintaining consistency in data analysis using VT. The researcher employed standardized procedures and settings in the TM to ensure the consistency of results across the entire CSSLR. The present study aims to discover answers to the research questions using all available data collection, creation and analysis methodologies with utmost care. Moreover, data analysis combined postulates from two theories to collect qualitative and quantitative data, and the results obtained through TM using VT are consistent even after repeated trials. Moreover, critics and students can independently analyse the textual data and cross-validate the findings to enhance the inter-rater reliability of the results of the current study by visiting VT. Therefore, this contributed towards the reliability of results and tools.

### **3.12 Chapter Summary**

The thematic investigation of *Sorrows of Sarasvati: The Lost River* aimed to uncover and interpret thematic elements portrayed in the narrative by employing both computational and qualitative thematic analysis. The research philosophy used in the study is post-positivism (post-empiricism), which acknowledges researchers' subjectivity while emphasizing the importance of empirical evidence and objectivity in the research process. The interpretative research paradigm emphasized the significance of thematic analysis based on contexts. An abductive research approach was applied in the study, which is top-down and bottom-up. It deals with both the validation as well as generation of theory. The mixed-method analysis combines quantitative analysis using text mining and qualitative analysis consisting of thematic analysis. The incorporation of both these methods required postulates from two (2) theories, which is known as OMT.

The time horizon of the study is cross-sectional. The validity of the research findings was ensured through careful data selection, calibration of Voyant Tools, and cross-validation of results. The reliability was emphasized by employing standardized procedures and encouraging inter-rater agreement. Ethical considerations guided the research process, ensuring proper citation and confidentiality. The study aimed to contribute to the broader field of DH and literary analysis, highlighting the potential of TM tools in uncovering underlying thematic patterns in literary works. The thematic investigation of CSSLR using VT is poised to offer valuable contributions to understanding the themes, enriching the comprehension of the human experience depicted in the novel.

## CHAPTER 4

### FINDINGS AND ANALYSIS

This Word document file of *Sorrows of Sarasvati: The Lost River* was uploaded on VT. It created the CSSLR, revealing five (5) panels on the main screen of VT (as shown in Figure No. 01). These five panels show five (5) different tools in the VT from a total of twenty-nine (29) tools. These five (5) panels include Cirrus/ word cloud visuals on the top-left side, the most frequent word of the Summary tool on the bottom-left side, collocates and their respective frequencies in the Collocates tool on the top-middle side, relative trends of words in the Trends tool on the top-right side, and contexts of words in Contexts tool on the bottom-right side. The quantitative findings constitute TM Knowledge patterns and nomenclature, while the qualitative data consists of emerging themes obtained through TM of VT.

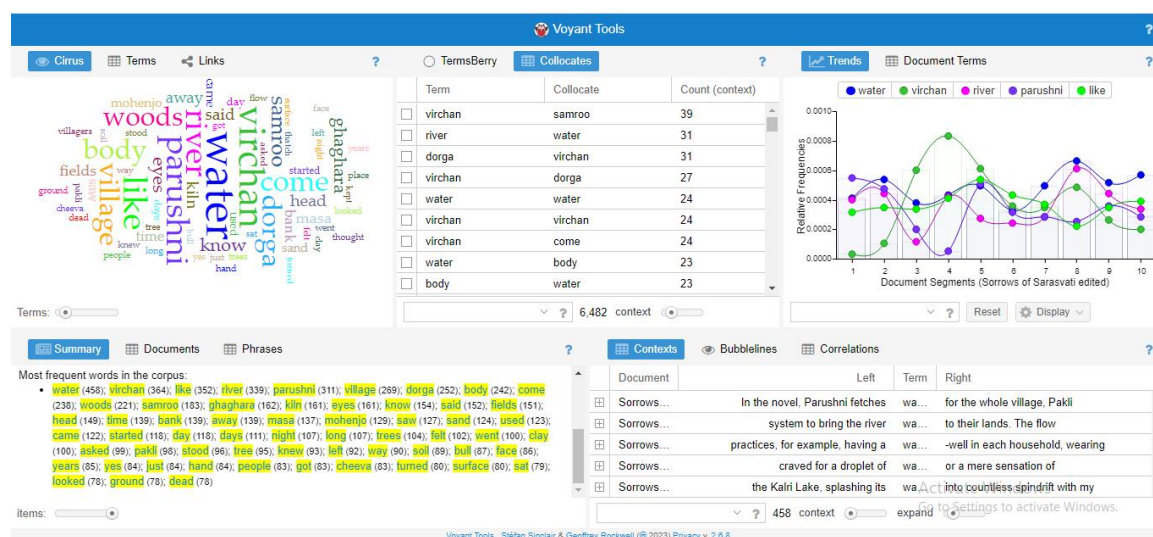


Figure 1. Main Window of Voyant Tools of CSSLR

#### 4.1 TM Knowledge Patterns

As one approach of previewing or distant reading, this investigation visualizes the CSSLR as a word cloud, revealing the important themes and prominent motifs throughout the novel. It generates a word cloud called Cirrus. Cirrus creates an intelligent and dynamic word cloud and a wide variety of themes, which may be looked up to 500 from 25. Voyant's Cirrus tool is superior to other word clouds including Wordle because the Cirrus tool reveals factual data analytical in nature of each topic and theme simply by putting the cursor on it, but the Wordle tool is unable to display quantifiable information.

In Cirrus, the most frequent words are broken down based on observable and quantifiable statistical information, and the word cloud that is provided displays both the presentation of words and the mining of themes. The following Cirrus/ word cloud visual consists of sixty-five (65) terms/words most frequent words:



Figure 2. Cirrus/ word cloud Visual

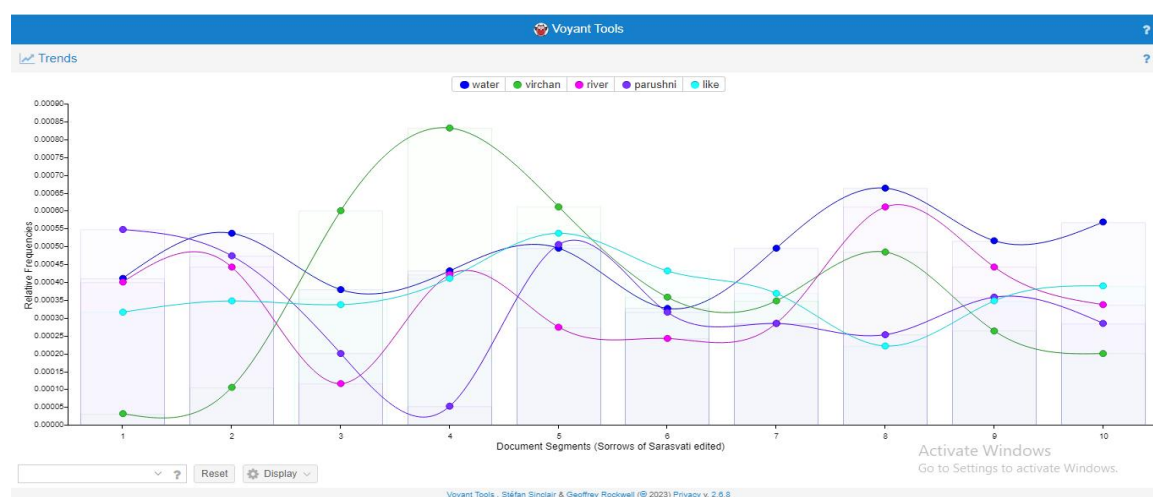
The summary tool reveals the most frequent words in the corpus along with their frequencies. The structured data is revealed here as significant terms from the corpus. In the present study, the CSSLR has one (1) document with 94,986 total words. The vocabulary density of the corpus is 0.074, with a total of 7,048 unique word forms. The readability index of the corpus is 8.306. There is an average of 18.6 words per sentence in the corpus. The corpus-based investigation of *Sorrows of Sarasvati: The Lost River* employed a range of commonly used words to explore different themes. This study obtained fifty-nine (59) most frequent words from the Summary tool, which was the focus of the current study. The most frequent words obtained from the Summary tool in the CSSLR are as follows:

water (458); virchan (364); like (352); river (339); parushni (311); village (269); dorga (252); body (242); come (238); woods (221); samroo (183); ghaghara (162); eyes (161); kiln (161); know (154); said (152); fields (151); head (149); away (139); bank (139); time (139); masa (137); mohenjo (129); saw (127); sand (124); used (123); came (122); day (118); started (118); days (111); long (107); night (107); trees (104); felt (102); clay (100); went (100); asked (99); pakli (98); stood (96); tree (95); knew (93); left (92); way (90); soil (89); bull (87); face (86); years (85); hand (84); just (84); yes (84); cheeva (83); got (83); people (83); surface (80); turned (80); sat (79); dead (78); ground (78); looked (78)



The Summary displays seven types of data in an organized format. The Summary offers a quick, literary description of the present corpus, ranging from word count, distinctive word count, vocabulary density, the average number of words in a sentence, most frequently occurring words, and exciting frequency points. The words in the Summary tool are yellow highlighted. One can reach the original corpus by clicking CTRL and the mouse-left click button. These words are separated by semicolons (;) making them distinguishable from other words and highlighting that each word portrays some significant themes in the corpus. The frequencies are shown in the small brackets (). These frequencies highlight the significance of these words in the corpus. The asterisk mark \* is used to show the allomorphs of the terms.

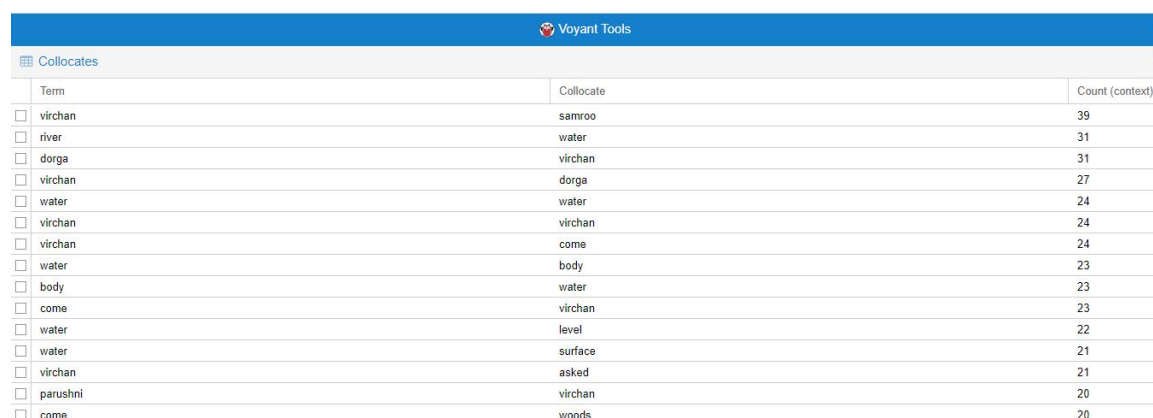
The graphs of the most frequent words are created through the Trends tool in the VT. These graphs allow one to observe the most common themes, significant variations, and core frequencies in different parts of the corpus. The CSSLR is shown in this graph as ten (10) separate segments, each containing one of the top five (5) most common terms. It enables viewers to establish the connection and recognize the association between the frequency of significant categories shown in ten (10) different segments. These words are shown with several combinations of colours, each representing the relative frequency of the term concerning its corresponding section. According to the Trends tool, the five (5) most common terms in the CSSLR are “[water](#) (458); [Virchan](#) (364); [like](#) (352); [river](#) (339); [Parushni](#) (311)”. These words appear more than any others. Figure No. 3 shows a graph that depicts the number of times each word appears in ten (10) segments.



*Figure 3. Terms Represented through Trends Tool in the Form of a Graph*

The Corpus Collocate Tool shows three (3) columns, known as term, collocate and count. Terms are the keywords to be collocated, Collocates are the terms discovered

around each keyword, and Count is the frequency of the collocates. For each of the ten (10) most common terms in the corpus, the most common collocates are displayed by default. In the Collocates Tool, one can add a keyword specification by entering a query term and pressing enter. Additionally, a slider controls the extent to which context should be considered while searching for collocates. The value determines how many words should be considered on either end of the keyword (creating a double-sized phrase window overall). The context setting is, by default, configured with five (5) words alongside each other. The slider has a value between 1 and 30. One can provide a list of stopwords one wants to eliminate by selecting the options icon (Figure No. 4).



Term	Collocate	Count (context)
<input type="checkbox"/> virchan	samroo	39
<input type="checkbox"/> river	water	31
<input type="checkbox"/> dorga	virchan	31
<input type="checkbox"/> virchan	dorga	27
<input type="checkbox"/> water	water	24
<input type="checkbox"/> virchan	virchan	24
<input type="checkbox"/> virchan	come	24
<input type="checkbox"/> water	body	23
<input type="checkbox"/> body	water	23
<input type="checkbox"/> come	virchan	23
<input type="checkbox"/> water	level	22
<input type="checkbox"/> water	surface	21
<input type="checkbox"/> virchan	asked	21
<input type="checkbox"/> parushni	virchan	20
<input type="checkbox"/> come	woods	20

Figure 4. Corpus Collocates Tool

## 4.2 Nomenclature

The nomenclature is the door to meanings (Fowler, 2014). It portrayed important information about the Proro-Dravidian people and their culture. The male characters **Virchan** (364), **Dorga** (252), **Samroo** (183), **Masa** (137), **Bull** (87), **Cheeva** (83), **Dhurva** (68), **Puran** (60), **Jhoorya** (8), **Geero** (7), **Sukra** (7), **Pindoo (Pundu)** (6), **Chamroo (Chamru)** (5), **Gujru** (5), **Gutta** (5), **Wango** (5), **Buti** (4), **Chandroo (Chandru)** (2), **Hunga** (2), **Mandra** (2), **Goga** (1), **Gutka** (1) and **Kola** (1), exhibit unique personalities and traits, contributing to their distinct characterization. Through the repeated appearance of their names, the writer establishes a sense of familiarity and prominence, allowing readers to form a connection with these characters. Whereas the female characters, particularly **Parushni** (311), **Pakli (Palki)** (98), **Gagri** (53), **Mati** (22), **Mayya** (9), **Kavasi** (5), **Jabbu** (4), **Nagri** (3), **Chuli** (1), **Harmi** (1), **Kari** (1), **Komi** (1) and **Lakhhi** (1), add depth and balance to the cast. Although appearing fewer times, their presence is equally significant where each character possesses its own individuality in shaping the story. **Zebu** (12) (the sacred bulls), **Pindroo** (11) (a brown deer), **bird** (56) and **peacock** 50 are also important in the

narration of the story. While **Parushni** (311) is the protagonist of the novel, **Virchan** (364) is used more frequently in the text and narration.

Various **places\*** (163) are mentioned in the novel, such as the **Kiln** (161) (a nesting place for **Pakli** (98) to bake pitchers, ewers etc). The unnamed village where these characters lived and **Mohenjo** (129) contribute to the rich tapestry of the story's setting. The repeated references to these locations evoke a sense of familiarity and establish a sense of connection. **Mohenjo** (129) is mentioned as a place that adds intrigue, drawing attention to this unique setting. Other places are also mentioned in the novel: **Hari Yupia** (23), **Mohenjo Daro (Mohenjo Daro)** (11), **Kalibangan** (9), **Vayarna** (9), **Chanhudaro** (2), **Banjoor** (2), **Miswal** (2), **Sowathi** (2), **Netan Dhawa** (1) and **Sapat Sindhu** (1). **Mana** (9) and **Maha Mayya** (9) are the supreme god and goddess in the story, respectively. **Yam** (20) is the god of death. Other mystical religious figures include **Bhikshu** (11), **Linga** (11), **Nanak (Nanaku)** (4), **Yakshina** (3), **Yakshini** (3), **Lingam** (2), **Shiv** (2) and **Shiva** (1). These ancient places and mystical figures connect to Proto-Dravidian times, and the lexicon indicates their meaning and symbolises their significance.

The term **Ghaghara** (162) serves as the name of a **river** (339), which is also known as **Sarasvati** (38). **Parushni** (311) is also the name of the **river** (339), which is called **Iravati** (1) as well (Parushni is also the name of a female character). Other rivers mentioned in the novel are **Sindhu** (64), **Drishadvati** (17), **Apaya** (6), **Shatudri** (4), **Asikni** (2) [also known as **Chandrabhaga** (1) or **Chandrabhagra** (1)], **Kubha** (2), **Vitasta** (2), **Arjikiya** (1) and **Vipasha** (1). **Kalri** (2) is the name of the **lake** (52) on the **river** (339) **Ghaghara** (162). In addition, **Chetar** (27), **Visakh** (8), **Poh** (9), **Magh** (8), **Phagun** (7), **Harh** (3) and **Phagan** (2) are the names of seasons and months. The plants and shreds famously mentioned are **Wheat** (64), **Katran** (12), **Silma** (11), **Chappri** (7), **Sesame** (7), **Tamarind** (6), **Saloo** (4), **Kapok** (3), **Kheip** (3), **Lana** (3), **Phog** (3), **Safflower** (3), **Vetchling** (3), **Ghorakh Pan** (2), **Barley** (2), **Dandi** (2), **Chhabri** (1), **Kasumba** (1), **Kheep** (1) and **Saffron** (1). Garments mentioned in the novel are **Lungi** (26), **Saloo** (4) and **Shawl** (1). **Kos** (11), **Karam** (10) and **Karo** (3) are the units of distance. Different animals and birds are also evident in the narration.

### 4.3 Theme of Nature and Wilderness

The theme of nature and wilderness is expressed through specific words in the text. The terms such as **water** (458), **village** (269), **woods** (221), **fields** (151), **bird\*** (121),

**trees** (104) and **tree** (95) create a vivid depiction of the natural environment. The word **woods** (221) instantly conjures images of a dense, mysterious forest, with towering **trees** (104) and gentle sunlight filtering through the leaves. It evokes a sense of adventure and exploration as if venturing into the heart of secrets of nature, as evident in the text as, **“Facing the sun, Parushni (311) entered the clump of trees (104) and made her way through their dense shade. Here, in the humid air, ... the rain used to pour down in torrents, and the air was quite damp there. Peepul and tamarind trees had grown in abundance”** (Awan & Khan, 2023, p. 25). Similarly, mentioning **fields** (151) brings vast, open spaces teeming with **life** (69) and an air of freedom. Whether the flourishing farmland and meadows are presented in the novel, **fields** (151) symbolize a connection to the earth and the vastness of nature.

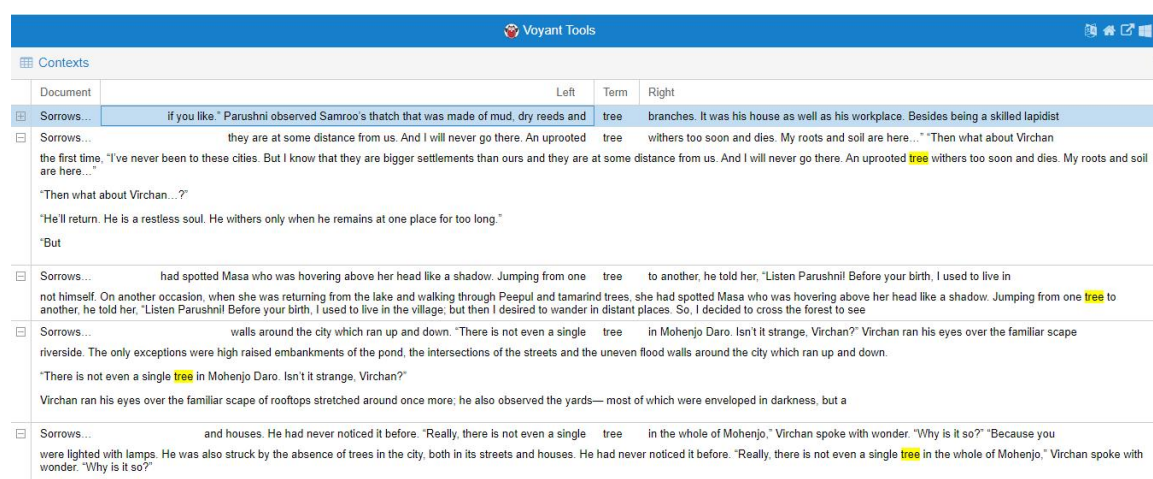
The most common word representing the theme of nature is water with the highest frequency of 458. In the examination of corpus via VT, the mention of **birds** (63) **“fluttering and twittering in delight”** suggests a lively and vibrant natural environment. The image of **birds** (63) **“splashing\* (37) water (458) with their wings (41)”** portrays an active interaction between the **birds** (63) and their surroundings. Furthermore, the repetition of the word **lake** (52) and the imagery of **sprinkling\* (7) water (458)** by **birds** (63) along the **lakeside** (2) creates a cascading effect, further emphasizing the abundance and connection of **water** (458) in the **village** (269) (Figure No. 5). The presence of **birds** (63) above the **lake** (52) signifies the interconnectedness of different elements within nature. This connection is evident in the text as, **“when the woods (221) were green with the dense foliage, sky covered with dark clouds, and the lake (52) brimming with water (458) and enticing the birds (63)”** (Awan & Khan, 2023, p. 363). Thus, the imagery of **birds** (63) and **water** (458) portrays the theme of nature in the novel.

Voyant Tools			
Contexts			
Document	Left	Term	Right
<input type="checkbox"/> Sorrows...	a mere illusion? I might, in fact, be flying above the Kalri Lake, splashing its	water	into countless spindrift with my wings. I might be ploughing its surface with my tail sprouting deep inside his body, pricking in veins and arteries like a thorny acacia. Is the sea of sand below a mere illusion? I might, in fact, be flying above the Kalri Lake, splashing its water into countless spindrift with my wings. I might be ploughing its surface with my tall feathers and absorbing its cool moisture into my plumage. My body is not an oven anymore but is chilled all
<input type="checkbox"/> Sorrows...	surrounded by flocks of birds; all of them are fluttering and twittering in delight, splashing	water	with their wings. And there are birds above the lake and above them, there is an oven anymore but is chilled all over, my mouth dribbles with luscious saliva. I am not alone, I am surrounded by flocks of birds; all of them are fluttering and twittering in delight, splashing water with their wings. And there are birds above the lake and above them, there is another lake of sprinkling water. We all flock to this land of soft seasons, bringing with us the briskness of
<input type="checkbox"/> Sorrows...	there are birds above the lake and above them, there is another lake of sprinkling	water	. We all flock to this land of soft seasons, bringing with us the briskness of surrounded by flocks of birds; all of them are fluttering and twittering in delight, splashing water with their wings. And there are birds above the lake and above them, there is another lake of sprinkling water. We all flock to this land of soft seasons, bringing with us the briskness of glaciers and gelid squalls; and we come every year. A bird is fluttering and chirping ahead of me, dabbling in
<input type="checkbox"/> Sorrows...	we come every year. A bird is fluttering and chirping ahead of me, dabbling in	water	. tweeting in utter delight, and I follow him. My tail consumes the briskness on the We all flock to this land of soft seasons, bringing with us the briskness of glaciers and gelid squalls; and we come every year. A bird is fluttering and chirping ahead of me, dabbling in water, tweeting in utter delight; and I follow him. My tail consumes the briskness on the surface of lake...
Is it the cool touch of water or the blaze of heat? Why am I slowing down			

Figure 5. The theme of Nature and Wilderness portrayed through Water and Birds

The inclusion of **trees** (104) and **tree** (95) (collectively **tree\*** (199)) amplifies the theme of nature and wilderness further, as these majestic entities embody strength,

resilience, and growth. Most of the time, the characters sit on the trees and talk, especially **Mamam Masa** (137) used to live in the trees. Moreover, **birds** (63) also flew through the **trees** (104) (Figure No. 6). The presence of “multiple species of **trees** (104)” suggests a thriving ecosystem abounding with diverse flora and fauna in nature, including **fish** (36). Intriguingly, the word **village** (269) appears amidst this natural imagery that contrasts the wilderness. While **village** (269) is typically associated with human habitation, it also signifies a harmonious coexistence between people and nature. However, “**there is not even a single tree\* (199) in Mohenjo Daro (11)**” (Figure No. 6), It suggests that these **trees\* (199)** are present at the **banks** (63) of the **river** (339) in almost all the villages or the “**dense woods**”. It suggests a community intimately intertwined with the surrounding **woods** (221), **fields** (151), and **trees\* (199)**, living in harmony with the natural world. Thus, the theme of nature is presented in connection to the habitat of human beings.



Document	Left	Term	Right
Sorrows...	if you like." Parushni observed Samroo's thatch that was made of mud, dry reeds and	tree	branches. It was his house as well as his workplace. Besides being a skilled lapidist
Sorrows...	they are at some distance from us. And I will never go there. An uprooted	tree	withers too soon and dies. My roots and soil are here..." Then what about Virchan
Sorrows...	the first time. "I've never been to these cities. But I know that they are bigger settlements than ours and they are at some distance from us. And I will never go there. An uprooted tree withers too soon and dies. My roots and soil are here..."		
Sorrows...	"Then what about Virchan...?"		
Sorrows...	"He'll return. He is a restless soul. He withers only when he remains at one place for too long."		
Sorrows...	"But		
Sorrows...	had spotted Masa who was hovering above her head like a shadow. Jumping from one	tree	to another, he told her, "Listen Parushni! Before your birth, I used to live in
Sorrows...	not himself. On another occasion, when she was returning from the lake and walking through Peepul and tamarind trees, she had spotted Masa who was hovering above her head like a shadow. Jumping from one tree to another, he told her, "Listen Parushni! Before your birth, I used to live in the village, but then I desired to wander in distant places. So, I decided to cross the forest to see		
Sorrows...	walls around the city which ran up and down. "There is not even a single	tree	in Mohenjo Daro. Isn't it strange, Virchan? Virchan ran his eyes over the familiar scape
Sorrows...	riverside. The only exceptions were high raised embankments of the pond, the intersections of the streets and the uneven flood walls around the city which ran up and down.		
Sorrows...	"There is not even a single tree in Mohenjo Daro. Isn't it strange, Virchan?"		
Sorrows...	Virchan ran his eyes over the familiar scape of rooftops stretched around once more; he also observed the yards—most of which were enveloped in darkness, but a		
Sorrows...	and houses. He had never noticed it before. "Really, there is not even a single	tree	in the whole of Mohenjo," Virchan spoke with wonder. "Why is it so?" "Because you
Sorrows...	were lighted with lamps. He was also struck by the absence of trees in the city, both in its streets and houses. He had never noticed it before. "Really, there is not even a single tree in the whole of Mohenjo," Virchan spoke with wonder. "Why is it so?"		

*Figure 6. The theme of Nature and Wilderness portrayed through Trees*

The word **woods** (221) also presents the theme of nature and wilderness. The mention of the **woods** (221) suggests a natural, untamed environment. It conjures up images of dense vegetation, trees, and wildlife, emphasizing the presence of nature in the surroundings. The mention of **bull** (87), **peacocks\*** (57), **bhookar\*** (21), **deer** (15), **porcupines\*** (15), **buffaloes\*** (5), **mongoose** (1), and **monkey** (1) further emphasizes the presence of wildlife in the **woods** (221), reinforcing the idea of a thriving ecosystem (Figure No. 7). Animal imagery is used to evoke emotions and produce a narrative in the text. Animal imagery beatifies the text and conveys complex ideas quickly. These animals also contribute to the theme of nature and wilderness in the novel. This theme of wilderness is present in the story where there is a marshland between the village and the woods, as evident in the text, “**A stretch of marshland lay beyond the edge of woods**” (Awan, & Khan, 2023, p. 26). They villagers “**felt blessed for having unique animals in**

their woods. Besides deer, mongoose and porcupine, woods had rare breeds of bulls and buffalos. They were wild and many villagers had fallen prey to their killing sprees. Luckily for them, they could not cross the marshland with their gigantic bulks and reach the village” (Awan, & Khan, 2023, pp. 57-58).

Voyant Tools			
Contexts			
Document	Left	Term	Right
<input type="checkbox"/>	Sorrows... They felt blessed for having unique animals in their woods. Besides deer, mongoose and porcupine, woods had rare breeds of bulls and buffalos. They were wild and many villagers had fallen that woods were dying. Thus, they were concerned about their own clump of trees around the drying lake and marshland. They felt blessed for having unique animals in their woods. Besides deer, mongoose and porcupine, woods had rare breeds of bulls and buffalos. They were wild and many villagers had fallen prey to their killing sprees. Luckily for them, they could not cross the marshland with their gigantic bulks and reach	woods	had rare breeds of bulls and buffalos. They were wild and many villagers had fallen
<input type="checkbox"/>	Sorrows... and nothing new appears; everything remains the same or gets replaced by something similar. The woods have always been inhabited since eternity by peacocks, deer, porcupines, and buffaloes. We, th... and her juices flowed down her thighs.	woods	have always been inhabited since eternity by peacocks, deer, porcupines, and buffaloes. We, th...
<p>There is this bewildering alteration of night and day everywhere. Nothing old vanishes and nothing new appears; everything remains the same or gets replaced by something similar. The woods have always been inhabited since eternity by peacocks, deer, porcupines, and buffaloes. We, the village dwellers, have been here since time immemorial, one generation replacing the other. There is no clue to how did this</p>			

Figure 7. Presence of Elements of Nature and Wilderness in Woods

There is a significant association of **woods** (221) with the **village** (269), **peacocks\*** (57), **desert\*** (57), and **lake** (52). The co-occurrence pattern of woods (221) with former terms is 14, 14, 11 and 8, respectively. The concordance of **woods** (221) with trees, birds, **animals** (56) and **air** (73) is 4 each. The graph of concordance of woods with those mentioned above eight (8) terms is shown in Figure No. 8. Both the **village** (269) and the **peacocks\*** (57) evoke a sense of nature and its untamed beauty. Consequently, when combined with the **village** (269) and the **peacocks\*** (57), the presence of **woods** (221) in the **village** (269) along with living beings effectively conveys a thematic representation of a natural setting, its habitat, its environment and its wilderness.

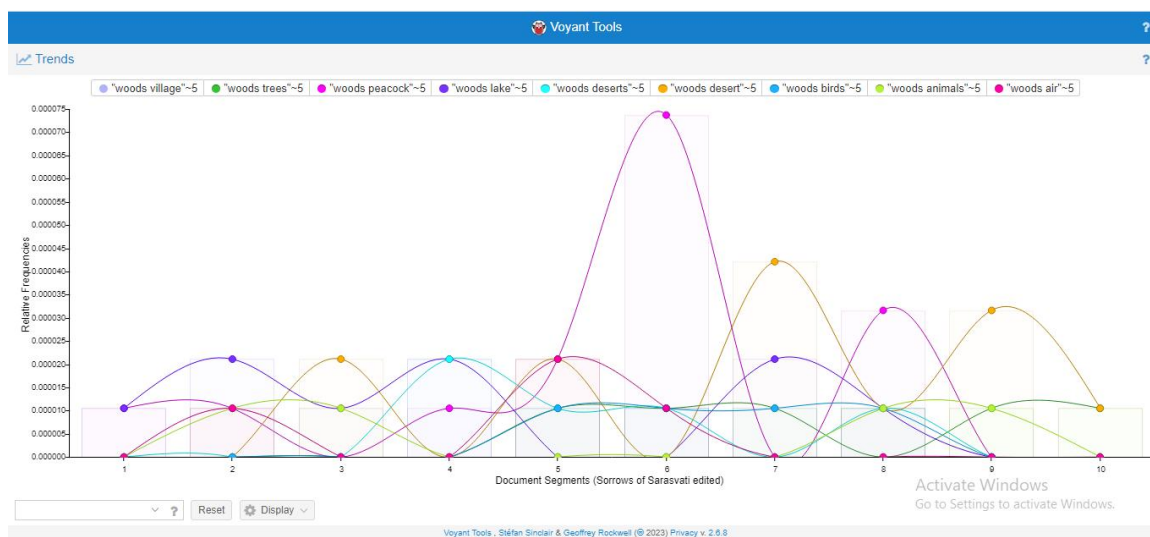


Figure 8. Concordance Graph of Woods with Elements of Nature

Figure No. 9 visually depicts the frequency trends pertaining to the concurrent coexistence of **woods** (221) with the **village** (269) and **peacocks\*** (57), thereby encapsulating the fundamental essence of the natural theme. By utilizing the Voyant trends tool, it can be observed that both the **village** (269) and **peacock** (50) exhibit an

equal co-occurrence with **woods** (221) in the initial four segments. However, a notable variation arises in the sixth segment, where the frequency of **peacock** (50) surpasses that of the **village** (269). The co-occurrence frequency is absent in the seventh and tenth segments. Within the figure, the green line corresponds to **peacock** (50) co-occurrence with **woods** (221). In contrast, the blue line represents the co-occurrence of the **village** (269) with **woods** (221). These concordance trends are higher in the sixth and eighth segments, contributing to the rising action in the novel.

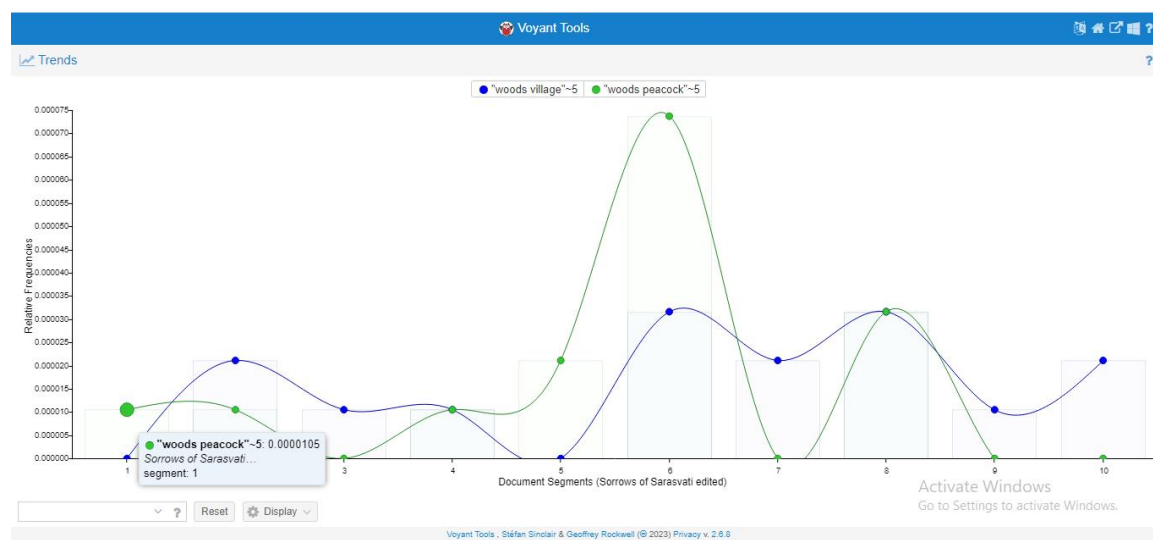


Figure 9. Relative Concordance of Woods with Village and Peacock

#### 4.4 Theme of Mystery and Fear of the Unknown

The high-frequency words such as **water** (458), **Virchan** (364), **river** (339), **Parushni** (311), **body** (242), **Samroo** (183), **eyes** (161), **bird\*** (121), **lake** (52) and **screamed** (50) portray the theme of mystery and unknown. The extract from VT portrays five (5) instances where the contextualization of **water** (458) shows that this high-frequency word portrays a theme of mystery and unknown (Figure No. 10). The given context of **water** (458) shows that as the **bird** (56) approaches a dense clump of trees enveloped in lush green foliage, he discovers a **lake** (52) nestled within. However, describing “its {**lake** (52)} **water** (458) did not sparkle” suggests an element of enigma and mystery. Moreover, there are “heap of dead birds” in he lake. Despite this, the **bird** (56) recognizes it as **water** (458), which evokes a sense of anticipation and intrigue. The bird’s physical condition, including its scorched brown wings, shrivelled beak, and inflamed **body** (242), intensifies the urgency and desperation for **water** (458).

The eternal questions filled the mind of **Parushni** (311) “Where is the opposite side of the river? ... How to reach there? ... What is all this? ... What is all this

**about? ... What is the reality of all this? ... Where is He if He exists? ... Why are we? ... How are we? ...**” (Awan & Khan, 2023, p. 376) poignantly capture the theme of mystery and the fear of the unknown. The protagonist’s mind, teeming with an array of existential questions, mirrors the human tendency to grapple with the vast unknowns of life. Questions like **“Where is the opposite side of the river?”** and **“What is the reality of all this?”** transcend their literal meaning, delving into the metaphorical journey across life’s uncharted waters. The river, in this context, becomes a symbol of life’s mysteries and uncertainties, with its opposite bank representing the elusive answers to life’s profound questions. Moreover, WH questions in the texts **“When (283) Who (249), What (223), Where (172), Why (160), How (148), Which (101), Whose (27) and Whom (16)”** are evident where characters want to know about the concept of god and goddess, their universe, their village and their existence.

The fear of the unknown is further heightened by the character’s reflections on Masa’s enigmatic inquiries about her life choices and the existence of Ghaghara, evident in the text as, **“Why did Masa smile while asking me why I lived in the village? ... Why did he ask me what if there were no Ghaghara, no water when I told him that I was living there because of Ghaghara? ...”** (Awan & Khan, 2023, p. 376). These questions challenge the protagonist’s understanding of her own existence, stirring a deep sense of unease and introspection. Masa’s smile, coupled with his thought-provoking questions, adds an element of ambiguity and mystery, compelling the protagonist to confront the unsettling reality that much of what we take for granted may be based on foundations as transient and unpredictable as a river. The relentless pursuit of these questions, while daunting, reflects the quintessential human quest to find meaning in the midst of the unknown, a journey fraught with fear yet driven by an unyielding desire for understanding.

The theme of mystery is also evident, particularly in the phrases, **“He spotted a lake (52) nestled in the middle of the clump, but its water (458) did not sparkle. Still, it was water (458)!”** (Figure No. 10) and, **“With his dwindling and sinking eyes (161), he saw his own image in every drop of water (458). There were faces of birds (63) who were ever born and who would ever be. They all resembled him, his look-alikes, and they craved for his arrival. Come down to us, we have been waiting only for you. Is it a mirage? A mirage? He asked every single bird (56), but none answered...”** (Figure No. 10). In these instances, it is evident that the water (458) of the lake (52) did



not sparkle which is not normal. Moreover, the mirage, the images of birds from the past and future, and the mystery connected to the **water** (458) make the theme more prominent.

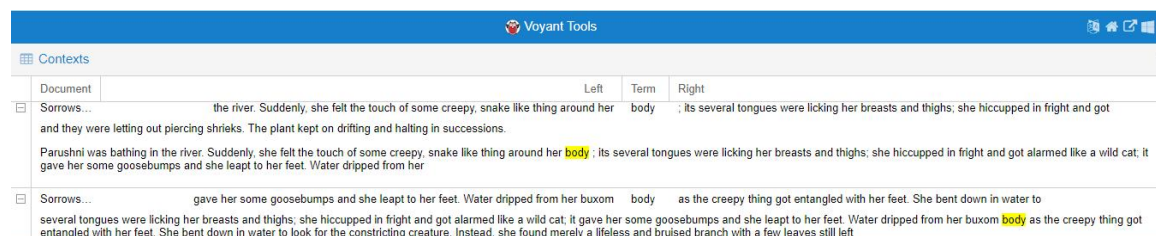
Voyant Tools			
Contexts			
Document	Left	Term	Right
☐ Sorrows...	his decent, he spotted a lake nestled in the middle of the clump, but its	water	did not sparkle. Still, it was water! He folded his claws and fluttered his wings
	of a bird returning to his nest. He was approaching a dense clump of trees overgrown with green foliage. In his decent, he spotted a lake nestled in the middle of the clump, but its <b>water</b> did not sparkle. Still, it was water! He folded his claws and fluttered his wings while landing on its surface. His scorched brown wings, palate-clung tongue, shriveled beak, and inflamed body were desperate for		
☐ Sorrows...	in the middle of the clump, but its water did not sparkle. Still, it was	water	! He folded his claws and fluttered his wings while landing on its surface. His scorched
	He was approaching a dense clump of trees overgrown with green foliage. In his decent, he spotted a lake nestled in the middle of the clump, but its water did not sparkle. Still, it was <b>water</b> ! He folded his claws and fluttered his wings while landing on its surface. His scorched brown wings, palate-clung tongue, shriveled beak, and inflamed body were desperate for water. With his dwindling and sinking eyes		
☐ Sorrows...	His scorched brown wings, palate-clung tongue, shriveled beak, and inflamed body were desperat...	water	. With his dwindling and sinking eyes, he saw his own image in every drop of
	did not sparkle. Still, it was water! He folded his claws and fluttered his wings while landing on its surface. His scorched brown wings, palate-clung tongue, shriveled beak, and inflamed body were desperate for <b>water</b> . With his dwindling and sinking eyes, he saw his own image in every drop of water. There were faces of birds who were ever born and who would ever be. They all resembled him, his		
☐ Sorrows...	With his dwindling and sinking eyes, he saw his own image in every drop of	water	. There were faces of birds who were ever born and who would ever be. They
	landing on its surface. His scorched brown wings, palate-clung tongue, shriveled beak, and inflamed body were desperate for water. With his dwindling and sinking eyes, he saw his own image in every drop of <b>water</b> . There were faces of birds who were ever born and who would ever be. They all resembled him, his look-alikes, and they craved for his arrival. Come down to us, we have been waiting		
☐ Sorrows...	picked him from the heap of dead birds in her one hand and, while dripping	water	into his open beak with the other, she said, "You too have come to the
	none answered...		
	No, it was not a mirage. He fell amidst a flock of fallen birds on the lakeside.		
	Parushni picked him from the heap of dead birds in her one hand and, while dripping <b>water</b> into his open beak with the other, she said, "You too have come to the lake to die?"		
	The bird wilted in her hand. He was dead.		
	"Meow, Meow," a peacock screamed from somewhere in		

Figure 10. The theme of Mystery and Unknown through Water, Bird and Body

The use of the phrase “**dwindling and sinking eyes** (161)” indicates the bird’s weakened state, adding to the air of uncertainty and mystery (Figure No. 10). As the **bird** (56) looks into the **water** (458), it sees its own reflection in each drop, but also the faces of **birds** (63) that have existed in the past and those that are yet to **come** (238). It suggests a mystical connection or a sense of the unknown, heightening the overall theme of mystery in the passage. In addition, the phrase “**Parushni** (311) **picked him from the heap of dead** (78) **birds** (63) **in her one hand and, while dripping water into his open beak with the other, she said, ‘You too have come to the lake** (52) **to die** (33)?’ The **bird** (56) **wilted in her hand** (84). **He was dead** (78)” (Figure No. 10). This suggests that there is something mysterious with the **water** (458) and **lake** (52) which makes the **birds** (63) **die** (33) in the water. Thus, the novel portrays mystery and unknown themes through **water** (458) and **Parushni** (311).

Further, Voyant collocate analysis shows that the second most common word that co-occurs with the **water** (458) is **body** (242) for 23 times (Figure No. 4). The collocation of **water** (458) with **body** (242) can be interpreted as representative of the theme of mystery and the unknown, particularly in the line, “**Parushni** (311) **saw a sturdy black** (48) **hand** (84) **approaching her from water** (458)” (Figure No. 12). **Water** (458), often associated with its fluid and elusive nature, carries connotations of depth, vastness, and hidden secrets. The human **body** (242), on the other hand, serves as a vessel of **life** (69) and consciousness, yet it also conceals intricate complexities that are not fully understood. When these two entities, **water** (458) and **body** (242), are brought together in collocation, they evoke a sense of enigma and intrigue, suggesting unexplored realms and unresolved

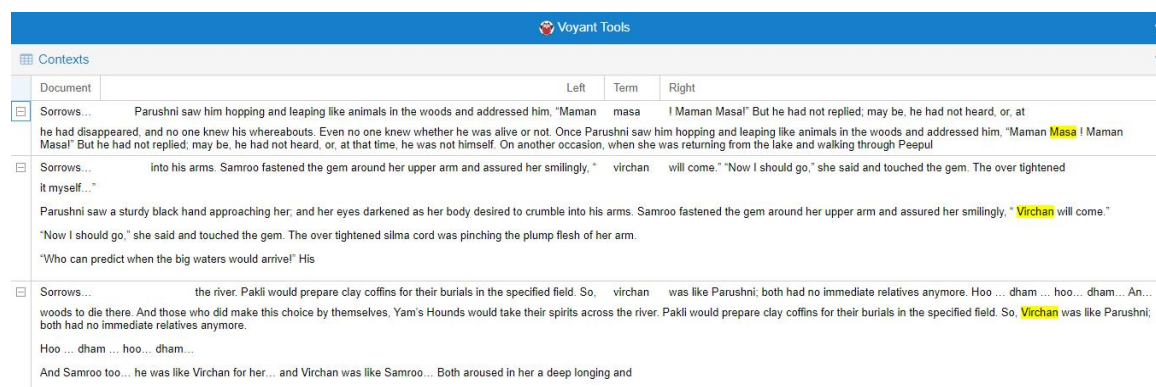
mysteries. The theme of mystery and unknown is also present in the instances: “**Parushni** (311) was bathing in the **river** (339). Suddenly, she felt the touch of some creepy, snake **like** (352) thing around her **body** (242); its several tongues were licking her breasts and thighs” (Figure No. 11), and “**Water** (458) dripped from her buxom **body** (242) as the creepy thing got entangled with her feet” (Figure No. 11).



Document	Left	Term	Right
Sorrows...	the river. Suddenly, she felt the touch of some creepy, snake like thing around her	body	; its several tongues were licking her breasts and thighs; she hiccupped in fright and got and they were letting out piercing shrieks. The plant kept on drifting and halting in successions.
Sorrows...	Parushni was bathing in the river. Suddenly, she felt the touch of some creepy, snake like thing around her	body	; its several tongues were licking her breasts and thighs; she hiccupped in fright and got alarmed like a wild cat. It gave her some goosebumps and she leapt to her feet. Water dripped from her

*Figure 11. Water, river, Parushni, and body revealing the unknown mystery*

The contextualization of the characters of **Virchan** (364), **Parushni** (311), **Dorga** (252) and **Masa** (137) also portray the theme of mystery and unknown. In the novel, the statement “So, **Virchan** (364) was **like** (352) **Parushni** (311); both had no immediate relatives anymore” and “**{Masa** (137)} disappeared, and no one knew his whereabouts” implies a sense of enigma surrounding the background and familial connections of **Virchan** (364), **Parushni** (311) and **Masa** (137). The absence of immediate relatives adds an air of mystery and raises questions about **Virchan’s** (32) origins and past. The repetition of the onomatopoeic sounds “**Hoo... dham... hoo... dham...**” and “**Meow** (53)” also create an atmosphere of suspense and intrigue (Figure No. 12). These sounds evoke a sense of the unknown as if hinting at hidden secrets or unexplored territories associated with **Virchan** (364) and **Masa** (137). Moreover, the narration of events also employed the theme of mystery and unknown, particularly evident through “**Parushni** (311) saw a sturdy **black** (48) **hand** (84) approaching her from **water** (458); and her **eyes** (161) darkened as her **body** (242) desired to crumble into his arms” (Figure No. 12) and “to be lost... to starve and **die** (33)... to drown” (Figure No. 28).



Document	Left	Term	Right
Sorrows...	Parushni saw him hopping and leaping like animals in the woods and addressed him, "Maman	masa	! Maman Masa!" But he had not replied; may be, he had not heard, or, at he had disappeared, and no one knew his whereabouts. Even no one knew whether he was alive or not. Once Parushni saw him hopping and leaping like animals in the woods and addressed him, "Maman Masa! But he had not replied, may be, he had not heard, or, at that time, he was not himself. On another occasion, when she was returning from the lake and walking through Peepul
Sorrows...	into his arms. Samroo fastened the gem around her upper arm and assured her smilingly, " virchan	will come."	"Now I should go," she said and touched the gem. The over tightened it myself..."
Sorrows...	Parushni saw a sturdy black hand approaching her; and her eyes darkened as her body desired to crumble into his arms. Samroo fastened the gem around her upper arm and assured her smilingly, " Virchan will come."		"Now I should go," she said and touched the gem. The over tightened silma cord was pinching the plump flesh of her arm. "Who can predict when the big waters would arrive!" His
Sorrows...	the river. Pakli would prepare clay coffins for their burials in the specified field. So, virchan	was like Parushni; both had no immediate relatives anymore. Hoo... dham ... hoo... dham... Ah... woods to die there. And those who did make this choice by themselves. Yam's Hounds would take their spirits across the river. Pakli would prepare clay coffins for their burials in the specified field. So, Virchan was like Parushni; both had no immediate relatives anymore.	
Sorrows...	Hoo ... dham ... hoo... dham...		And Samroo too... he was like Virchan for her... and Virchan was like Samroo... Both aroused in her a deep longing and

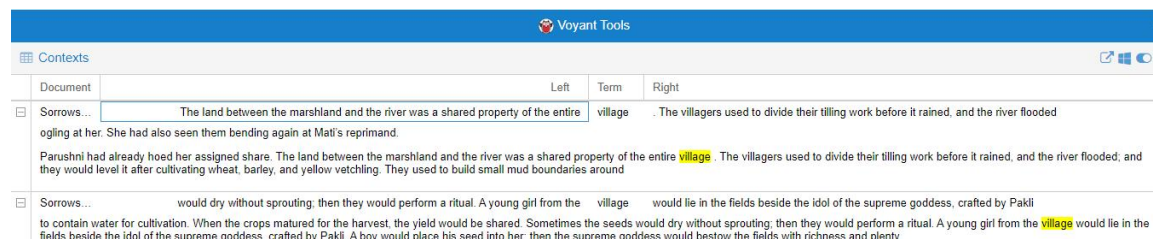
*Figure 12. Theme of Mystery and Unknown with Masa, Virchan and Parushni*

## 4.5 Theme of Superstitions

The theme of superstitions is also evident in *Sorrows of Sarasvati: The Lost River*, where different characters exhibit this theme. An evident example of superstition is, **“A young girl from the village (269) would lie in the fields (151) beside the idol of the supreme goddess, crafted by Pakli (98). A boy would place his seed into her; then the supreme goddess would bestow the fields (151) with richness and plenty”** (Figure No. 13). This statement portrays a fertility ritual rooted in superstition. In this practice, a young girl from the village (269) lies beside an idol of the supreme goddess, and a boy places his seed into her. This ritual is conducted with the belief that it will invoke the goddess’s blessings, bestowing the fields (151) with fertility and abundance. This practice underscores the superstitious beliefs prevalent in the community, where the act of procreation is intricately linked to supernatural forces. It reflects a belief system where rituals and offerings to deities are perceived as essential for agricultural prosperity. This theme of superstition is tied to the idea of nature worship and the reliance on rituals to control natural phenomena, highlighting the characters’ deep-rooted beliefs in the supernatural and superstitions for prosperity. Thus, the theme of the supernatural is evident, as it is practised by the villagers for the prosperity of their fields in everyday life.

In one other instance, it is stated in the novel, **“He crossed to the opposite bank (139) where spirits of the dead (78) were believed to reside”** (Figure No. 13). This instance alludes to another superstition which is the belief in spirits of the dead (78) residing on the opposite bank (139). Crossing to the bank (139) where spirits are believed to reside suggests a superstitious fear or reverence for the spirit world. This belief reflects the characters’ apprehension about the supernatural realm and their desire to avoid any potential malevolent influence from the spirits. It adds a layer of mysticism and superstition to the narrative, illustrating how the characters navigate their lives while being mindful of these supernatural forces. This theme of superstition also underscores the characters’ connection to their ancestral beliefs and the mysterious elements that shape their world view. Both these instances highlight the theme of superstition as a significant aspect of the characters’ lives and the cultural context of the novel. Superstitions shape their actions, decisions, and perceptions of the world around them, adding depth to the narrative and emphasizing the influence of the supernatural on their daily lives. The novel is has many instances of superstitions- one instance of superstition includes **“She knew that after leaving human body, a spirit would transmigrate into**

bulls and peacocks; afterwards, they take it to the other side of the river, the abode of the spirits” (Awan & Khan, 2023, p. 36).

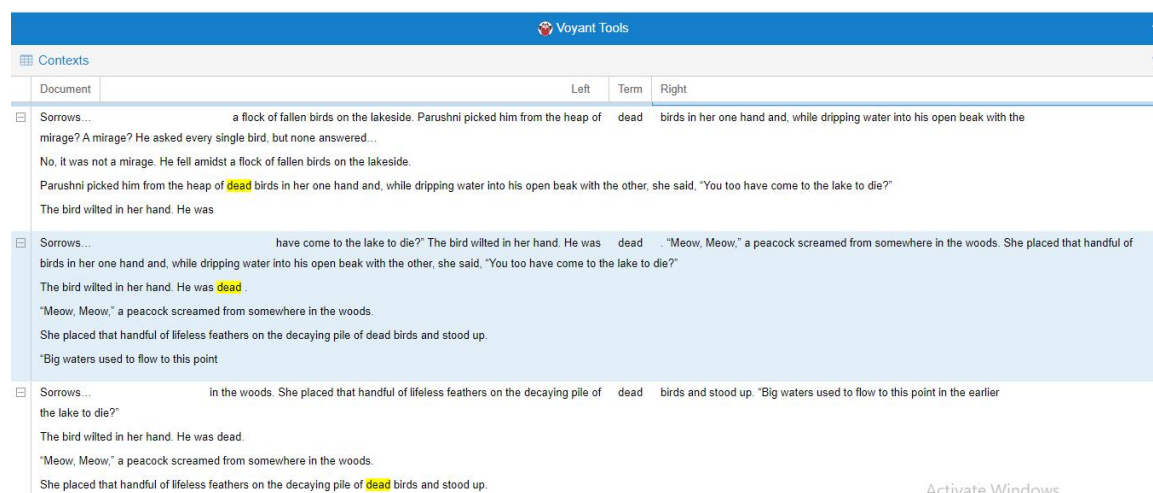


Document	Left	Term	Right
Sorrows...	The land between the marshland and the river was a shared property of the entire	village	. The villagers used to divide their tilling work before it rained, and the river flooded
Sorrows...	would dry without sprouting; then they would perform a ritual. A young girl from the	village	would lie in the fields beside the idol of the supreme goddess, crafted by Pakli

Figure 13. Theme of Hope and Superstitions

## 4.6 Theme of Death

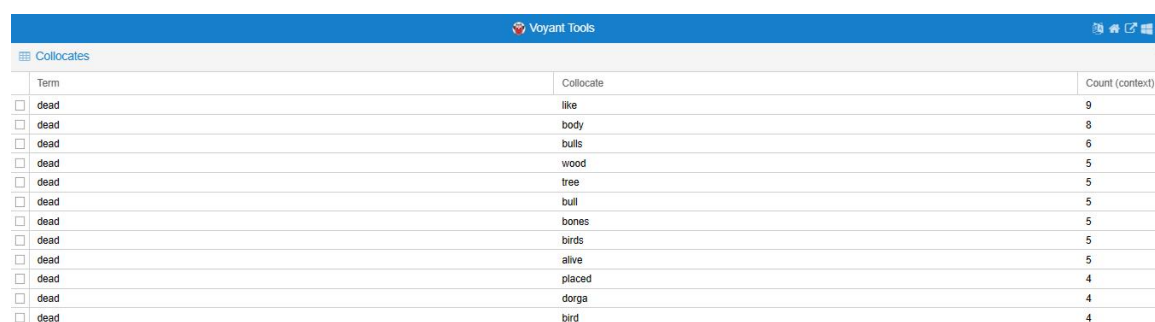
The repetition of the terms **dead** (78), **die** (33), **died** (24), **death** (19), **kill\*** (15), **deaths** (2) and **deadly** (1) in the corpus emphasizes its significance within the narrative and its contribution to the theme of death. In the contextualization of the Contexts tool, the word **dead** (78) contributes to the theme of death by highlighting the presence of deceased beings and emphasizing their lifeless state. The phrase “**heap of dead (78) birds\*** (121)” creates a vivid image of a collection of deceased avian creatures, symbolizing mortality and the fragility of **life** (69) (Figure No. 14). **Parushni’s** (66) action of picking up a **bird\*** (121) from this heap reinforces the theme of death, as she interacts with a lifeless **body** (242). Moreover, the use of **dead** (78) to describe the **bird\*** (121) in **Parushni’s** (66) **hand** (84) emphasizes its lack of vitality and confirms its lifeless condition. This word choice underscores the theme of death, leaving no room for ambiguity about the bird’s state. The phrase “**decaying pile of dead (78) birds\* (121)**” also enhances the theme of death by stressing the deterioration and decomposition of the deceased creatures (Figure No. 14).



Document	Left	Term	Right
Sorrows...	a flock of fallen birds on the lakeside. Parushni picked him from the heap of	dead	birds in her one hand and, while dripping water into his open beak with the
Sorrows...	have come to the lake to die? The bird wilted in her hand. He was	dead	“Meow, Meow,” a peacock screamed from somewhere in the woods. She placed that handful of
Sorrows...	in the woods. She placed that handful of lifeless feathers on the decaying pile of	dead	birds and stood up. “Big waters used to flow to this point in the earlier

Figure 14. The theme of Death through the Terms Died and Die

The repetition of **leave\*** (103), **kiln** (161), **dead** (78) and **die** (33) emphasizes the prevalence and significance of death within the narrative. The Voyant Tool collocate analysis further shows that the word **dead** (78) often collocates with the words **like** (352), **body** (242), **bulls** (59), **woods** (221), **tree** (95), **bones** (41), **birds** (63) and **alive** (27). This collocation adds to the theme of death in the corpus in several ways (Figure No. 15). First, the collocation with **like** (352) suggests a comparison between something lifeless and the concept of death, emphasizing the similarity or resemblance between the two concepts (Figure No. 15). This conveys a sense of the inanimate nature associated with death. Additionally, “Beyond **Pakli’s** (54) **kiln** (161), there was a **field\*** (151) for burial; the deceased were buried in **clay** (100) coffins” also contributes to the theme of death and ultimate journey (Figure No. 31). It signifies the intimate connection between life and death, where the same clay that nurtured life through pottery in Pakli’s kiln is used to craft coffins for the deceased. This juxtaposition of creation and burial underscores the cyclical nature of existence and the inevitability of death. The use of clay coffins also reflects the characters’ reverence for the earth and their belief in a spiritual connection with the land, highlighting the theme of death and its profound significance in their lives.



Term	Collocate	Count (context)
<input type="checkbox"/> dead	like	9
<input type="checkbox"/> dead	body	8
<input type="checkbox"/> dead	bulls	6
<input type="checkbox"/> dead	wood	5
<input type="checkbox"/> dead	tree	5
<input type="checkbox"/> dead	bull	5
<input type="checkbox"/> dead	bones	5
<input type="checkbox"/> dead	birds	5
<input type="checkbox"/> dead	alive	5
<input type="checkbox"/> dead	placed	4
<input type="checkbox"/> dead	dorga	4
<input type="checkbox"/> dead	bird	4

*Figure 15. Collocates of Dead*

The collocations with **body** (242), **bulls** (59), **woods** (221), **tree** (95), **bones** (41) and **birds** (63) further emphasize the connection between the **dead** (78) and various elements associated with mortality. These terms symbolize different aspects of death, such as decomposition, lifelessness, decay, and the passage of time. Additionally, the collocation of **alive** (27) contrasts starkly, highlighting the juxtaposition between **life** (69) and death. This pairing reinforces the theme of mortality by emphasizing the absence of vitality in the context of death. The most frequently occurring words with the word **dead** (78) show that these occurrences enrich the theme of death by providing diverse perspectives and associations that evoke a sense of finality, stillness, and the sombre nature of the subject matter (Figure No. 15).

The collocation with the **body** (242) directly links the **dead** (78) to the physical form that is devoid of **life** (69), underscoring the cessation of vital functions and the absence of **life** (69) force. There are other instances of death in the novel, particularly, “she said, ‘You too have come to the **lake** (52) to **die** (33)?’ The **bird** (56) wilted in her **hand** (84). He was **dead** (78)” (Figure No. 10) and “to stave and **die** (33) the driest of the deserts” (Figure No. 28). This statement encapsulates the theme of death in the novel, highlighting the stark inevitability and presence of mortality. The dialogue about coming to the lake to die, coupled with the bird’s demise, symbolizes the inescapable cycle of life and death, mirroring the broader theme of impermanence and human fragility prevalent in *Sorrows of Sarasvati: The Lost River*. The collocations of the **dead** (78) provide nuanced insights into the theme of death. These associations evoke imagery and connections: **body** (242) implies human mortality, **bulls** (59) suggest the cycle of life and death in nature, **woods** (221) and **tree** (95) signify the inevitable decay of life, while **bones** (41) and **birds** (63) symbolize the transient nature of existence, collectively enriching the exploration of death within the text. The connection of the **dead** (78) is evident in the Cirrus visual presented in Figure No. 16.

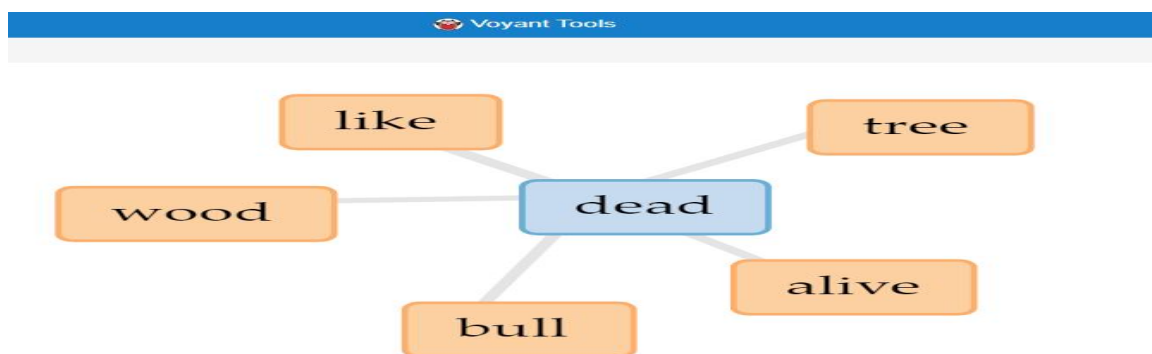


Figure 16. Link of Dead with its Collocates

By highlighting the contextual relationships between words and illuminating the semantic nuances present in a text, collocations help us comprehend the themes of a corpus (Biber & Reppen, 2015). When used together, these word pairs or groupings frequently have a particular connotation or share a close semantic relationship. The contextual importance of certain terms within the corpus is revealed by locating and examining collocations (Bennett, 2010). For example, in the context of *Sorrows of Sarasvati: The Lost River*, collocates of the **dead** (78) suggest a thematic connection, allowing for a deeper exploration of the theme of death within the text. It is evident from the graphic representation that these collocates appear more often in the last two segments of the novel which suggests that the theme of death is more evident and prominent in

these two last segments of the novel (Figure No. 17), which is also evident in the text as, “I was as sure to **kill** (15) the **bird** (56) as I am about the fact that one day my soul would be taken away by Yam’s Hounds, leaving this worthless **body** (242) to be buried in some earthen vessel. I could see the spectre of **death** (19) riding the flapping wings of bhookar” (Awan & Khan, 2023, p. 76), “He could not guess for how long the old man was **dead** (78). He took his **body** (242) to the **woods** (221) to place it in one of the hollow **trees\*** (199) where there were **bones** (41) of his own parents” (Awan & Khan, 2023, p. 358), “The black bull was **dead** (78)” (Awan & Khan, 2023, p. 363); and “and again you’ll **kill** (15) him... It’ll continue for ever” (Awan & Khan, 2023, p. 364).

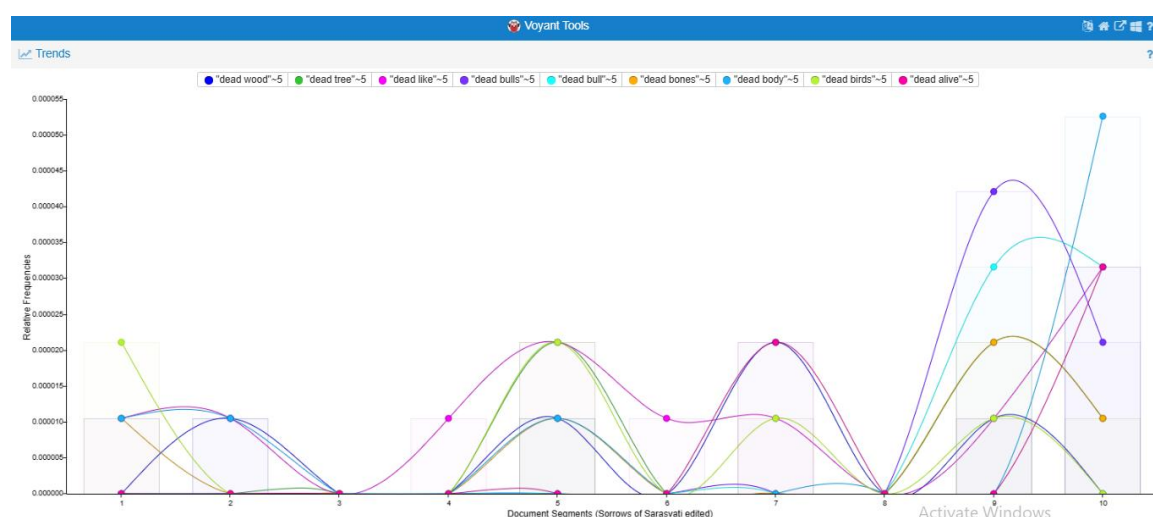


Figure 17. Graph of Collocation of Dead through Trends Tool

#### 4.7 Theme of Community and Shared Responsibility

**Fields** (151) and **village** (269) contribute to the theme of community and shared responsibility by symbolizing a shared resource that requires collective effort and cooperation. The statements, “**This field will yield mere weed; then, you will have to harvest and eat it**” and “**It was Parushni’s communal duty to fetch water for all the households in the village as they had divided daily chores among them...**” (Awan & Khan, 2023, p. 58) suggest a sense of community and shared responsibility for the field’s outcome. It implies that if the **fields** (151) produce only weeds, everyone in the community will be affected and have to endure the consequences. The Corpus contextualization of the **fields** (151) shows questioning the state of the field where the speaker emphasizes the importance of working together and taking responsibility for the cultivation and productivity of **land** (51) in the **village** (269).

The **fields** (151) in the fictional work *Sorrows of Sarasvati: The Lost River* become a symbol of the community's shared livelihood, sustenance, and well-being. Furthermore, the various contexts of the **fields** (151) indicate that the well-being of the community depends on the success or failure of the **fields** (151). It implies that the community members must actively nurture and maintain the **fields** (151) to ensure their productivity and the collective benefit it brings. The shared responsibility fosters a sense of unity and mutual support among the community members. Overall, **fields** (151) contributes to the theme of community and shared responsibility by representing a common resource that requires the collaboration and collective effort of the community members. It is evident that **"The land between the marshland and the **river** (339) was a shared property of the entire **village** (269). The villagers used to divide their tilling work before it rained"** (Figure No. 13).

**Village** (269) and its residents also contribute to the theme of community and shared responsibility. The shared property of the **land** (51) and the collective efforts of the villagers in dividing the tilling work, building mud boundaries, and sharing the harvest highlight the theme of community and shared responsibility, as highlighted in the text, **"They used to build small mud boundaries around **fields** (151) to contain **water** (458) for cultivation. When the crops matured for the harvest, the yield would be shared"** (Figure No. 13). It showcases the interconnectedness and cooperation among the villagers in their agricultural endeavours, emphasizing the importance of unity and collective action for the well-being of the entire community. It underscores the interdependence and shared fate of the community, emphasizing the need for active involvement in sustaining the **fields** (151) for the benefit of all inhabitants of the **village** (269).

Every character has a profession as well or there is some work allocated to each person in the village. For instance, Parushni fetches water for the whole village, Pakli bakes earthenware, Gagri is an expert in hunting animals and birds, and Virchan and Samroo plough the lands. Samroo, like Pakli, is also an artist who knows how to make pictures on stone, gold, silver and sea-shells. He also carves farming instruments out of stones. Samroo's house was his workplace. **"Besides being a skilled lapidist, he used to carve stamp seals and draw sketches on conches"** (Awan & Khan, 2023, p. 50). Further, **"It was Parushni's communal duty to fetch water for all the households in the village as they had divided daily chores among them"** (Awan & Khan, 2023, p. 58). Thus, there is a concept of community and shared responsibility in the novel.



## 4.8 Theme of Loss and Sorrow

The entire narrative of the novel revolves around the themes of loss and sorrow. Various high and low-frequency words **water** (458), **like** (352), **river** (339), **Parushni** (311), **village** (269), **Dorga** (252), **Ghaghara** (162), **fields** (151) and **sorrow** (35) contribute to this theme. In the corpus, **village** (269) presents the theme of loss and sorrow through its gradual transformation and eventual decay. In the extract in Figure No. 18, the phrase **“it is inching towards its extinction”** emphasizes the theme of loss and sorrow. The word extinction conveys a sense of finality and irrevocable decline, indicating that the **village** (269) is facing an inevitable fate of disappearing or ceasing to exist. This notion evokes a profound sense of **sorrow** (35) as the community and its way of **life** (69) are on the brink of being lost forever. The **village** (269), therefore, becomes a metaphor for the transience of **life** (69) and the **sorrow** (35) that accompanies the loss of what was once cherished and believed to be eternal. Moreover, **“Dorga (252) noticed tears (20) falling down her cheeks and falling off the child (30). Nobody in the boat (57) knew (93) why they were crossing onto the other side”** contributes to the theme of loss and sorrow (Figure No. 28).

Document	Left	Term	Right
Sorrows...	the closing dham-dham rising from Parushni's thatch. The paradisiacal village on the bank of	ghaghara	, which Parushni believed was impervious to change, gradually succumbs to the corrosive powers of t...
of novel is informed by a faintly tragic note: the omniscience of death and decay from the opening sequence to the closing dham-dham rising from Parushni's thatch. The paradisiacal village on the bank of Ghaghara , which Parushni believed was impervious to change, gradually succumbs to the corrosive powers of time and the changing course of the river.			
<b>Fictionalizing History, Historicizing Fiction</b>			
The geography of Sorrows of Sarasvati is a palimpsest			
Sorrows...	through the story of a small town located on the banks of the legendary river	ghaghara	, known as Sarswati in the Arian scriptures, the Rigveda and the Mahabharata. Tarar's work was
these areas. Tarar has done his research and has, therefore, created an authentic historical fiction about the Indus Valley civilization through the story of a small town located on the banks of the legendary river Ghaghara , known as Sarswati in the Arian scriptures, the Rigveda and the Mahabharata. Tarar's work was more daunting than any other writer of that lost world. In Sorrows of Sarasvati, tracing pre-Buddhist and pre-Arian			
Sorrows...	Sorrows of Sarasvati is the story of one such village on the banks of river	ghaghara	. It is inching towards its extinction, since, with every passing year, Ghaghara, the source of
local Dravidians. However, still some villages, away from the main city centres like Harappa and Mohenjo-Daro, existed in isolation. Sorrows of Sarasvati is the story of one such village on the banks of river Ghaghara . It is inching towards its extinction, since, with every passing year, Ghaghara, the source of all land fertility and life in that region, is drying up as its tributary rivers change their course. According to			
Sorrows...	banks of river Ghaghara. It is inching towards its extinction, since, with every passing year,	ghaghara	, the source of all land fertility and life in that region, is drying up as
like Harappa and Mohenjo-Daro, existed in isolation. Sorrows of Sarasvati is the story of one such village on the banks of river Ghaghara. It is inching towards its extinction, since, with every passing year, Ghaghara , the source of all land fertility and life in that region, is drying up as its tributary rivers change their course. According to the available archaeological research, the legend of a lost river in the			
Sorrows...	the blue dome of sky seemed to meet a gigantic bulk of water. It was	ghaghara	, one of the mighty rivers. She descended the mound on the other side, stepping carefully
long distance she had covered to reach there, and she mounted the mound feeling afresh. She could now behold that the blue dome of sky seemed to meet a gigantic bulk of water. It was Ghaghara , one of the mighty rivers. She descended the mound on the other side, stepping carefully on the layers of hot pebbles, and reached the riverbank. She tried to balance the ewer on the sloping ground			

Figure 18. The theme of Loss and Sorrow through River, Village and Ghaghara

The collocation analysis reveals a consistent association between the term **village** (269) and **river** (339), highlighting the vital dependence of the fertility of the **village** (269) on the **Ghaghara** (162) **river** (339) and its **water** (458). The significant occurrence of **village** (269) and **river** (339) co-occurring 16 times emphasizes the pivotal role of the **river** (339) and its **water** (458) in sustaining the prosperity of the **village** (269). However, **“the water (458) levels of the river (339) gradually decline over time”**, resulting in its eventual scarcity (Figure No. 18). This scarcity evokes a prevailing theme of loss and sorrow within the community. In addition, the frequency trend using the Trends tool

reveals a significant association between the terms **village** (269) and **river** (339). This association between these two words is particularly prominent in the first and sixth segments, where their co-occurrence frequency peaks. Conversely, the fifth and eighth segments exhibit the lowest occurrence of this association (Figure No. 19). It means that the first and sixth segments of the corpus portray the themes of loss and sorrow in one way or another where gradual decline in the water levels of the river are evident.

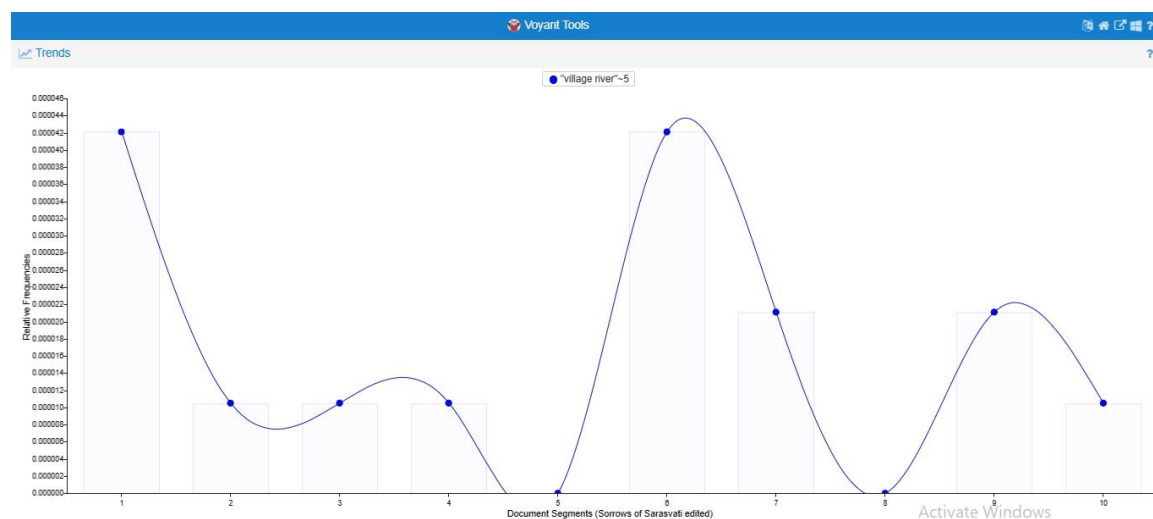


Figure 19. Graph of Collocation between Village and River

**Ghaghara** (162) also contributes to the theme of loss and sorrow in several ways. The mention of the **Ghaghara** (162) **river** (339) in the novel serves as a symbolic representation of loss and sorrow within the context of the paradisaical **village** (269). Initially, the **village** (269) is portrayed as a haven untouched by the passage of time and impervious to change, as perceived by **Parushni** (311). However, as the narrative unfolds, it becomes evident that the **village** (269) undergoes a gradual transformation, marked by the relentless erosion caused by the changing course of the **Ghaghara** (162) **river** (339). This transformation embodies the theme of loss and sorrow, as the **village** (269) once considered idyllic, succumbs to the corrosive forces of time and nature. The shifting course of the **river** (339) serves as a metaphor for the unpredictable and uncontrollable nature of change, highlighting the inevitable loss and sorrow accompanying such transformations. Moreover, “**inching towards its extinction**” also refers to the theme of loss and sorrow as it refers towards the extinction of the **village** (269) (Figure No. 18).

The Voyant Collocates tool analysis reveals a significant collocation pattern of **Ghaghara** (162) with its river **bank** (139), **water** (458), **village** (269) and **river** (339) supporting their representation of the theme of loss and sorrow. The repeated occurrence of this collocation 27, 18, 8 and 7 times, respectively, underscores the profound

association of **Ghaghara** (162) with the **bank** (139), **water** (458), **village** (269) and **river** (339), highlighting their inseparable connection. However, an ironic twist unfolds as the depletion of **water** (458) from the **river** (339) **water** (458) leads to the gradual drying of **Ghaghara** (162), resulting in the overall loss of the village's fertility. This loss is particularly poignant considering the previous dependence of the **village** (269) on **water** (458) for sustenance and prosperity, thus highlighting the theme of loss and sorrow. The varying co-occurrence frequencies of these terms reflect the shifting dynamics and associations surrounding loss and sorrow throughout the text. This observation highlights the thematic significance of the concurrent frequency of **Ghaghara** (162) with its collocates in elucidating the portrayal of loss and sorrow within the corpus.

In the Voyant Trends tool context, the collocation between **Ghaghara** (162) and **water** (458) is absent in segments 2, 3, 4, 7, 8 and 9. There is low collocation between **Ghaghara** (162) and **village** (269) in segments 2, 3, 4, 5 and 9, completely missing in segments 8 and 10. Similarly, there is low collocation between **Ghaghara** (162) and **river** (339) in segments 2, 3 and 8, and this collocation is completely missing in segments 5 and 10. Moreover, there is low collocation between **Ghaghara** (162) and **bank** (139) in segments 1 and 2, which is completely missing in segments 5. All these most frequent collocates of **Ghaghara** (162) have their maximum frequency in the sixth segment (Figure No. 20). Thus, the sixth segment contributes to the theme of loss and sorrow within the analysed corpus. The theme of loss and sorrow is absent or is present in low frequency in other segments of the novel.

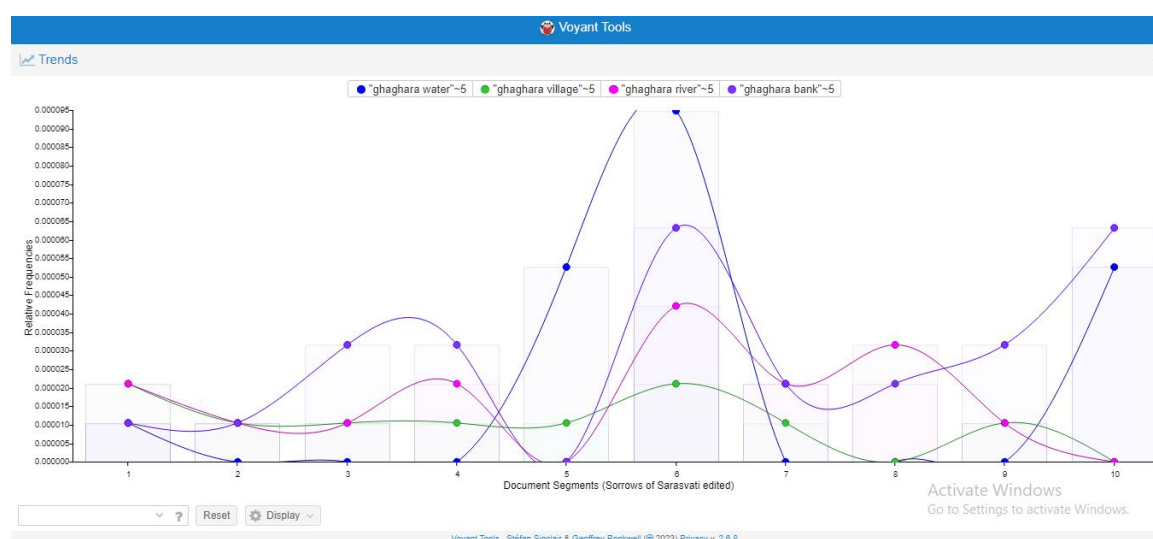


Figure 20. Graph of Collocates of Ghaghara with Water, Village, River and Bank

**Parushni** (311) also significantly reinforces the theme of loss and sorrow within *Sorrows of Sarasvati: The Lost River*. First, **Parushni** (311) firmly believed that village (269) was impervious to change and would remain unblemished by decay. However, her beliefs were shattered when the **village** (269) began to deteriorate, exposing the fallacy of her convictions. This contradiction highlights the theme of loss and sorrow as **Parushni** (311) experiences the disappointment and sadness of witnessing the village's transformation. Second, **Parushni** (311) had the responsibility of fetching **water** (458) for the entire **village** (269). However, with the depletion of the **river** (339), not only were the villagers deprived of a vital resource, but **Parushni** (311) herself lost her daily routine (Figure No. 21). Therefore, the firm beliefs of **Parushni** (311) about the **village** (269), and its association to elements of nature constitutes the theme of loss of sorrow.

The scarcity of **water** (458) because of the depletion of the **river** (339) not only deepens the theme of loss and sorrow for the **village** (269) as a whole but also intensifies it for **Parushni** (311) individually. The disruption of her customary tasks accentuates the sorrow as she mourns the loss of a significant aspect of her life (69) and the way of the **village** (269). In a nutshell, the beliefs of **Parushni** (311) being proven wrong and the depletion of the **river** (339) contribute to the theme of loss and sorrow. The contrast between her initial conviction and the decay of the **village** (269) and the disruption of her routine due to the diminishing water (458) supply underlines the overarching theme of loss and sorrow throughout the narrative. The phrases “**bad omen**”, “**fallen bird**”, “**peacock scream**,” and “**dense shade**” also portray the theme of loss and sorrow, and their association with the character of **Parushni** (311) in the text suggests that the theme is connected to the life of characters as well (Figure No. 21).

Document	Left	Term	Right
☐ Sorrows...	dham-dham rising from Parushni's thatch. The paradisiacal village on the bank of Ghaghara, which is informed by a faintly tragic note: the omniscience of death and decay from the opening sequence to the closing dham-dham rising from Parushni's thatch. The paradisiacal village on the bank of Ghaghara, which believed was impervious to change, gradually succumbs to the corrosive powers of time and the changing course of the river.	parushni	believed was impervious to change, gradually succumbs to the corrosive powers of time and the changing course of the river.
Fictionalizing History, Historicizing Fiction The geography of <i>Sorrows of Sarasvati</i> is a palimpsest of many			
☐ Sorrows...	for thousands of years before the arrival of the Arians as invaders. In the novel, the present-day Cholistan desert is most likely associated with Saraswati. The Dravidian cultures that flourished in those lands survived for thousands of years before the arrival of the Arians as invaders. In the novel, fetches water for the whole village, Pakli bakes earthenware, Gagri is expert in hunting animals and birds, Virchan and Samroo plough the lands. Samroo, like Pakli, is also an artist who knows how to make	parushni	fetches water for the whole village, Pakli bakes earthenware, Gagri is expert in hunting animals and birds, Virchan and Samroo plough the lands. Samroo, like Pakli, is also an artist who knows how to make
☐ Sorrows...	was not a mirage. He fell amidst a flock of fallen birds on the lakeside. waiting only for you. Is it a mirage? A mirage? He asked every single bird, but none answered... No, it was not a mirage. He fell amidst a flock of fallen birds on the lakeside. <b>Parushni</b> picked him from the heap of dead birds in her one hand and, while dripping water into his open beak with the other, she said, "You too have come to the lake to die?" The	parushni	picked him from the heap of dead birds in her one hand and, while dripping
☐ Sorrows...	while engrossed in observing the fallen bird. "Meow, Meow," the peacock screamed. Facing the s... because she had heard from the elders that it was a bad omen. But that day she did it unintentionally while engrossed in observing the fallen bird. "Meow, Meow," the peacock screamed. Facing the sun, <b>Parushni</b> entered the clump of trees and made her way through their dense shade. Here, in the humid air, her sweat began to trickle from her neck down to her back. The rain used to pour	parushni	entered the clump of trees and made her way through their dense shade. Here, in the humid air, her sweat began to trickle from her neck down to her back. The rain used to pour

Figure 21. Association of Theme of Loss and Sorrow with the Character of Parushni

The Voyant Trends tool allows for a comprehensive understanding of the varying co-occurrence patterns, providing valuable insights into the representation and significance of **Parushni** (311) in relation to different aspects that present the theme of loss and sorrow at a broader level. This corpus-driven study demonstrates that the contextualization of frequently occurring words such as **water** (458), **river** (339), **Parushni** (311), **village** (269), **Ghaghara** (162), **fields** (151), **dead** (78) and **life** (69), along with their collocates and varying frequency trends collectively contribute to the overarching theme of loss and sorrow. Through careful analysis, it becomes evident that these words play a significant role in shaping the narrative's exploration of emotional pain and its profound impact on various aspects of **life** (69).

Furthermore, **Parushni** (311) often collocates with **Virchan** (364), **river** (339), and **Samroo** (183) to exhibit the theme of loss and sorrow with the frequencies 20, 15 and 14, respectively. The analysis conducted using the Voyant frequency trends tool also provides comprehensive insights into the varying co-occurrence frequencies associated with **Parushni** (311). Upon examining the visual representation, it becomes apparent that **Parushni** (311) and **Samroo** (183) co-occurrence is most prominent in the ninth segment. On the other hand, **Parushni** (311) and **Virchan** (364) reach their highest frequency in the eighth segment. At the same time, **Parushni** (311) and **river** (339) exhibit their peak occurrences in the second and ninth segments. These findings shed light on the distribution and prominence of the various associations of **Parushni** (311) throughout the corpus. These collocations underscore the interconnectedness of characters and the narrative's emotional depth, emphasizing how the characters' experiences are intertwined with loss and sorrow, enriching the thematic exploration within the novel.

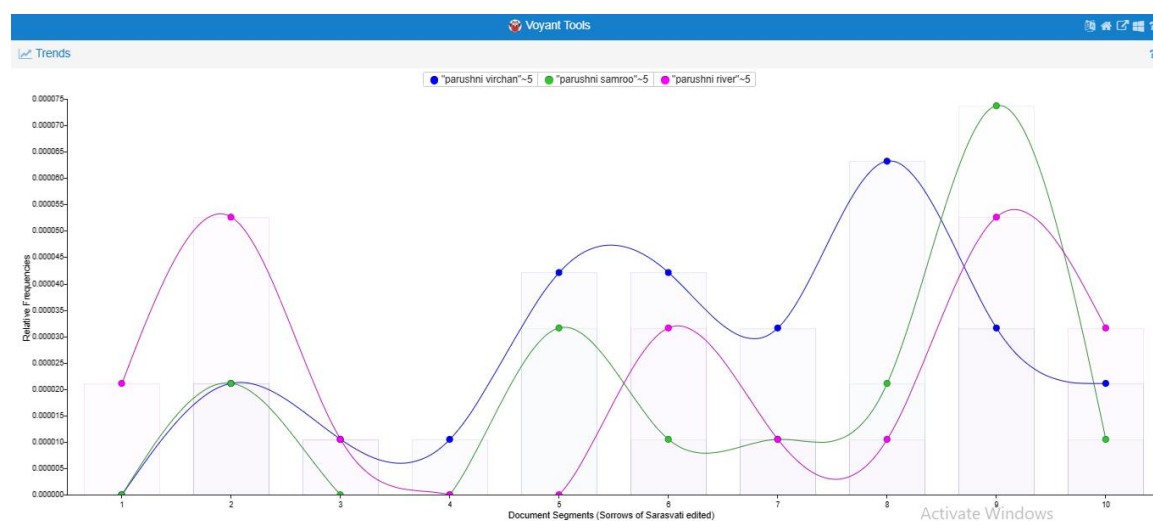


Figure 22. Collocations of Parushni with Virchan, River, and Samroo

## 4.9 Theme of Emotions, Sensations and Nostalgia

The words **felt** (102), **like** (352), **Dorga** (252), **feel\*** (171), **dead** (78) and **tears** (20) convey a range of emotions, sensations, and nostalgia. These carefully chosen words establish a deep sense of connection between the reader and the avian characters, allowing us to **feel\*** (171) their experiences. **Felt** (102) emphasizes the **birds** (63) ability to experience emotions and creates a palpable sense of their inner world. It is evident in the phrases “**Parushni** (311) **felt** (102) a **peacock\*** (57) **screaming\*** (50) inside her” and “**she felt** (102) a **strange foreboding**” that **felt** (102) contributes to the overall emotions and nostalgia. This theme of emotions underscores the profound emotional lives that **birds** (63) possess, challenging the notion that they are solely creatures of instinct (Figure No. 23). Moreover, “**Dorga** (252) noticed **tears** (20) falling down her cheeks and falling of the **child** (30),” contributes to the theme of emotions (Figure No. 28).

Voyant Tools			
Contexts			
Document	Left	Term	Right
☐ Sorrows...	few lines on the belly of peacock; there appeared a human face. "Meow, Meow," Parushni	felt	a peacock screaming inside her. She knew that after leaving human body, a spirit would softened it by chewing with her sparse teeth. Then, she dipped it daintily into the pigment container and drew a few lines on the belly of peacock; there appeared a human face.
	"Meow, Meow," Parushni	felt	a peacock screaming inside her. She knew that after leaving human body, a spirit would transmigrate into bulls and peacocks; afterwards, they take it to the other side of the river, the abode of the
☐ Sorrows...	of the bushy stretch. The land beneath her feet gently rose into a mound. She	felt	as if the sky was leaning nearer. She bent and, grasping the ewer, threw a she crossed the tall reeds, patches of sandy land became visible through the thick bushes. She stopped at the edge of the bushy stretch. The land beneath her feet gently rose into a mound. She felt as if the sky was leaning nearer. She bent and, grasping the ewer, threw a clod of clay across the mound with full force. She stood still, without even breathing, till she heard a splash
☐ Sorrows...	tightly wrapped lungi also. As she ran her hand on her well-rounded hips, she	felt	her flesh had a raised scar or weal as if some venomous insect had crawled faced difficulty in movement. Parushni, too, was well endowed and had to cover hers. She loosened the strips of her tightly wrapped lungi also. As she ran her hand on her well-rounded hips, she felt her flesh had a raised scar or weal as if some venomous insect had crawled and stung there last night. She unwrapped her lungi, placed it on the potshards beside the ewer, and stood straight
☐ Sorrows...	bulging holms amid the riverbed; it appeared as a scarcely visible line of mounds. She	felt	a strange foreboding. Peacocks and bulls carry our disembodied souls to the opposite bank. The keep on flowing? Her mind was riddled with questions. Her gaze shifted towards the other bank that lied beyond the bulging holms amid the riverbed; it appeared as a scarcely visible line of mounds. She felt a strange foreboding. Peacocks and bulls carry our disembodied souls to the opposite bank.
The peacock of the woods screamed out again but from somewhere inside her still-breathing body that was now shivering with			

Figure 23. Felt portraying the Theme of Emotions, Sensations, and Nostalgia

The **fields** (151) mentioned in the fictional novel portray various themes depending on the text where it has been utilized. It is connected to the theme of emotions, sensations and nostalgia, where the description, “**Clouds** (27) of **dust** (48) hung above the **fields** (151) were replaced by a **cover\*** (74) of **sand** (124)” signifies a physical transformation and degradation of the once cultivable **land** (51). This imagery emphasizes the loss of arable **land** (51) and the inability to continue agricultural practices, adding to the overall emotions and sensations and the theme of loss and sorrow. Furthermore, the mention of the **thatches\*** (130) being “**blown away** (139) by the **storm\*** (9)” implies the destruction and devastation caused by external forces. The loss of shelter and protection from the storm intensifies the sense of nostalgia and vulnerability experienced by the individuals relying on the **fields** (151) for their livelihoods. In addition, the phrases, “**Fields** (151) will **never** (126) be **green** (35) again... Let us **go** (197)”, and “we shall

be known only by the potsherds of the vessels we've left here” are also connected to the theme of emotions, sensations and nostalgia (Figure No. 24).

Document	Left	Term	Right
Sorrows...	covered with the sand, yards, streets, the riverbanks, shrubs. Clouds of dust hung above the	fields	were replaced by a cover of sand... Only a few patches of the soil were
Sorrows...	by the storm... The sky remained clear afterwards... It did not bring even a single drop of rain... Everything was covered with the sand, yards, streets, the riverbanks, shrubs. Clouds of dust hung above the	fields	were replaced by a cover of sand... Only a few patches of the soil were visible from underneath the thick sheet of sand... All the ditches around the village had been filled by the sand
Sorrows...	and sniffing. It was that time of the year when big waters would flood their	fields	. A baby died of hunger in the village. In the evening, when they were coming
Sorrows...	stuck like a branch... Almost all the animals had died, birds were also dying... But he was still there, bellowing and sniffing.		
Sorrows...	It was that time of the year when big waters would flood their	fields	. A baby died of hunger in the village. In the evening, when they were coming back after burying the baby and adding a stone to the pebbled path, every one of them was thinking about
Sorrows...	For how long it will be flowing? A few more months at the most... The	fields	have already been covered with sand and our thatches have been blown away by the
Sorrows...	her.		
Sorrows...	"You think that I've lost my mind," she could read their minds.		
Sorrows...	"No... but... The river is drying up... For how long it will be flowing? A few more months at the most... The	fields	have already been covered with sand and our thatches have been blown away by the storm... Our wells are also drying up... Everything is drawing closer to the end and you say that we will
Sorrows...	embers, "it is never the end of everything... If we have half handful of wheat,	fields	will be green again... Put it back into the pitcher." Samroo did so and handed
Sorrows...	go, we shall be known only by the potsherds of the vessels we've left here;	fields	will never be green again... Let us go," Virchan insisted. "No..." this time she said
Sorrows...	full of foliage paintings over her luscious body. Yet even that skeletal Parushni was dear to him.		
Sorrows...	"Parushni, let us go, we shall be known only by the potsherds of the vessels we've left here;	fields	will never be green again... Let us go." Virchan insisted.
Sorrows...	"No..." this time she said it in a way that he could not ask her again.		

Figure 24. Fields portraying the Theme of Emotions, Sensations, and Nostalgia

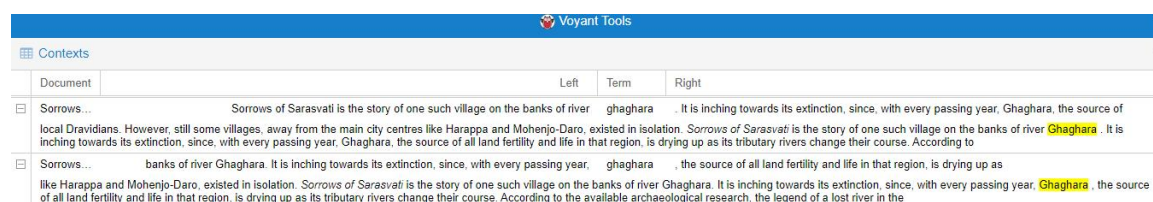
Additionally, the frequent use of the word **like** (352) invites the reader to imagine the **birds'** (121) sensations. It draws parallels between their experiences and the experiences of **people** (83) living in the **village** (269). This comparison evokes a sense of shared humanity, fostering a connection between us and the avian protagonists. Furthermore, the words **dead** (78) and **die** (33) evoke a nostalgic tone and serve as a poignant reminder of the fleeting nature of **life** (69). The phrase “**he was walking\* (101) dead (78) on his feet (59)**”, elicits a bitter-sweet sentiment, as the **bird's\*** (121) gestures, laden with emotions and sensations, are ultimately transient and subject to the passage of time. This theme of nostalgia adds depth to the narrative, prompting reflection on the impermanence of experiences and the inherent beauty of fleeting moments.

#### 4.10 Theme of Natural Transformations and Environmental Changes

The theme of natural transformations and environmental changes is also evident in the fictional masterpiece. **Ghaghara** (162) **river** (339) presents the theme of natural transformations and environmental changes. **Ghaghara** (162) contributes to the theme of transformations in nature by symbolizing the changing course of a **river** (339) and its impact on the **village** (269) and the surrounding region. The **Ghaghara** (162) **river** (339), described as the “**source of all land fertility and life**”, is undergoing a significant transformation as its tributary rivers change their course, and “**Ghaghara (162) is drying up**” (Figure No. 25). This transformation of the **river** (339) implies a larger process of change and loss in the natural environment. The theme of natural transformations along with environmental changes is depicted through the drying up of the **Ghaghara** (162) **river**

(339), which directly affects the **village** (269) and its inhabitants. Moreover, the **kiln** (161) contributes to the theme of transformations in nature and environmental change as well, particularly evident in the lines, “**the smoke from the **kiln** (161) was soaring higher**” (Figure No. 31). This all resulted in desertification of the village as evident, “**Sand was encroaching rapidly and turning the woods into a desert**” (Awan & Khan, 2023, p. 332), and “**The sand has encroached upon your tree... Let us go to the inner part of the woods; some trees are still alive there**” (Awan & Khan, 2023, p. 362).

The **river** (339), once a lifeline for the community, is gradually losing its life-sustaining properties and becoming a source of sorrow. This transformation signifies the vulnerability and fragility of nature and the interconnectedness of different elements within an ecosystem. The word **Ghaghara** (162) serves as a symbol of this transformation, representing the changing dynamics of the natural world, and “**Ghaghara (162) is inching towards its extinction**” (Figure No. 25). It highlights the theme of how natural elements can undergo significant alterations over time, impacting the lives of individuals and communities reliant on them. As the **Ghaghara** (162) **river** (339) undergoes a drying-up process, and its tributary rivers alter their course, the impacts extend beyond the **river** (339) alone. The **bank** (139), **water** (458), **village** (269), and other rivers connected to **Ghaghara** (162) are also affected, experiencing their own natural transformations in response to the changing environmental conditions.



Document	Left	Term	Right
<input type="checkbox"/>	Sorrows... Sorrows of Sarasvati is the story of one such village on the banks of river	ghaghara	. It is inching towards its extinction, since, with every passing year, Ghaghara, the source of local Dravidians. However, still some villages, away from the main city centres like Harappa and Mohenjo-Daro, existed in isolation. Sorrows of Sarasvati is the story of one such village on the banks of river Ghaghara. It is inching towards its extinction, since, with every passing year, Ghaghara, the source of all land fertility and life in that region, is drying up as its tributary rivers change their course. According to
<input type="checkbox"/>	Sorrows... banks of river Ghaghara. It is inching towards its extinction, since, with every passing year,	ghaghara	. the source of all land fertility and life in that region, is drying up as like Harappa and Mohenjo-Daro, existed in isolation. Sorrows of Sarasvati is the story of one such village on the banks of river Ghaghara. It is inching towards its extinction, since, with every passing year, Ghaghara, the source of all land fertility and life in that region, is drying up as its tributary rivers change their course. According to the available archaeological research, the legend of a lost river in the

Figure 25. The theme of Natural Transformations through Ghaghara

The word **Ghaghara** (162) frequently collocates with the **bank** (139), **water** (458), **village** (269) and **river** (339) for 27, 18, 8 and 7 times, respectively. It portrays the theme of natural transformation and environmental change. These closely linked collocations of **Ghaghara** (162) are also subject to the effects of environmental changes, thus undergoing their own natural transformations. By frequently appearing alongside **Ghaghara** (162), words like **bank** (139), **water** (458), **village** (269) and **river** (339) establish a strong connection to the **Ghaghara** (162) **river** (339) itself. These collocations have a symbiotic relationship between the **Ghaghara** (162) **river** (339) and the elements it interacts with within its surroundings. The changes occurring in the **Ghaghara** (162) **river** (339) have ripple effects on these associated components, amplifying the natural transformation.



The frequent collocations of **Ghaghara** (162) with the **bank** (139), **water** (458), **village** (269) and **river** (339) strengthen the portrayal of natural transformations in the given context. These interconnected elements, influenced by the environmental changes affecting the **Ghaghara** (162) **river** (339), contribute to the theme of natural transformations. The frequency trends associated with the simultaneous presence of **Ghaghara** (162) with the **bank** (139), **water** (458), **village** (269) and **river** (339) capture the essence of the natural transformations and environmental change (Figure No. 26). The drying river prompts reflection on time, impermanence, and environmental stewardship. Overall, the comprehensive theme reveals the intricate relationship between human existence, the natural environment, and the dimensions of life, emphasizing the importance of water as a resource, the transformational power of rivers, and the consequences of environmental changes in ecosystems. The presence of these collocations highlights the interdependence of these elements within the ecosystem.

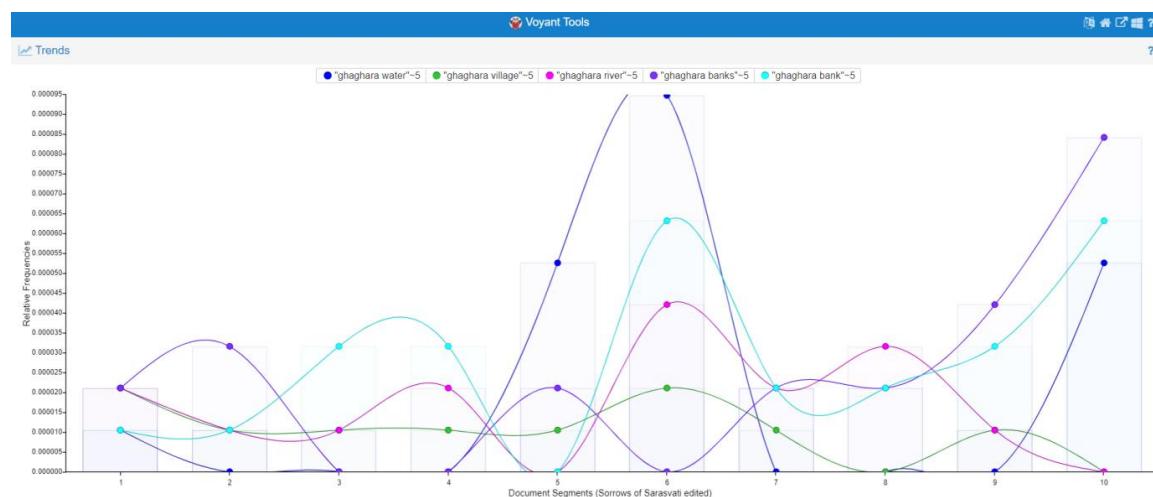


Figure 26. Collocations of Ghaghara Showing Theme of Environmental Change

#### 4.11 Theme of Impermanence and Transience

The writer adeptly employs the words **time** (139), **years** (85), **kiln** (161), **away** (139), and names of months and seasons to weave a theme of impermanence and transience throughout the narrative. The term **time** (139) acts as a constant presence, underscoring the relentless march of moments and the ephemeral nature of our experiences. Through the repeated use of **time** (139), the writer invites contemplation on the brevity of **life's\*** (88) joys and sorrows, urging readers to cherish the present and embrace the ever-changing nature of our surroundings. Furthermore, with the word **years** (85), the writer emphasizes the passage of time and the fleeting nature of existence. Each mention of **years** (85) serves as a reminder of the temporal framework within which

the story unfolds, highlighting the inevitability of change and the impermanence of all things. The “**potsherds and ashes**” from the **kiln** (161) also contribute towards the theme (Figure No. 31). The word **away** (139) shows the movement of things that do not last for a long time in a place (Figure No. 27). Additionally, the names of seasons and months **Chetar** (27), **Visakh** (8), **Poh** (9), **Magh** (8), **Phagun** (7), **Harh** (3) and **Phagan** (2) show the theme of impermanence and transience. The progression of the **day** (118), **days** (111) and **night** (107) also exhibits the theme of impermanence and transience.

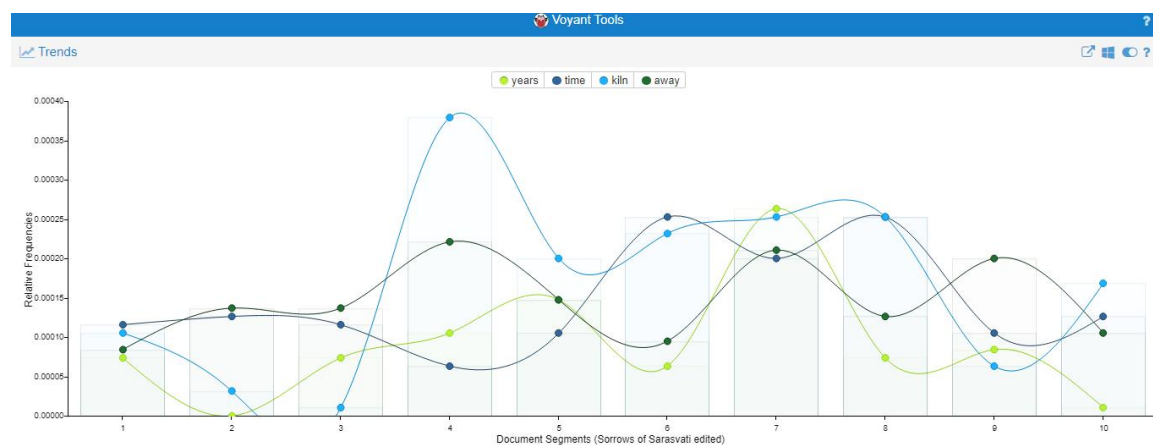


Figure 27. Trends Tool Graph of Theme of Impermanence and Transience

## 4.12 Theme of Journey

The repeated use of **Dorga** (252), **started** (118), **leave\*** (103), **went** (100), **way** (90), **turned** (80), **came** (122) and **long** (107) exhibit the theme of journey. These carefully chosen words evoke a sense of movement, progress and distance. The word **started** (118) indicates the beginning of the journey, signalling a moment of embarkation and setting the narrative in motion. It captures the anticipation and excitement that often accompanies the commencement of a voyage. The theme of journey is present in “**Dorga** (252) was **leaving\*** (103)” (Figure No. 28), which has the same meaning as **started** (118) and **went** (100). With the **way** (90), the writer emphasizes the path or route taken during the journey. This word invites readers to envision the physical and metaphorical roads travelled by the characters, highlighting the importance of navigation and direction.

Document	Left	Term	Right
Sorrows...	the flooding rivers? Who in a sane mind would leave the safety of Mohejo? ... But	dorga	was leaving. It was the first body of water he had seen, and it was to be lost in the dark woods, to starve and die the driest of the deserts, and to drown in the flooding rivers? Who in a sane mind would leave the safety of Mohejo? ... But Dorga was leaving. It was the first body of water he had seen, and it was the first boat he had boarded.
Sorrows...	child was drowsy and his lips, stiffened by cold, were unable to grip the nipples.	dorga	noticed tears rolling down her cheeks and falling on the child. Nobody in the boat lying without moving a muscle in his mother's lap. She was trying to suckle him on her freezing breast. The child was drowsy and his lips, stiffened by cold, were unable to grip the nipples. Dorga noticed tears rolling down her cheeks and falling on the child. Nobody in the boat knew why they were crossing onto the other side. A bit farther from her, another passenger sat beside a huge

Figure 28. The theme of Journey through Dorga and Leaving

The mention of **turned** (80) suggests moments of decision or change in direction along the journey. This word conveys a sense of adaptability and the possibility of new paths unfolding before the characters. Moreover, **Came** (122) also shows the end of the journey and the arrival of the people. Lastly, **long** (107) emphasizes the distance and duration of the journey, evoking a sense of endurance and perseverance. It underscores the challenges faced and the transformational nature of the voyage itself. These words and their repetition construct a narrative framework that encapsulates the essence of a journey. They invite readers to follow along as the characters traverse physical, emotional, or metaphorical landscapes, emphasizing the transformational power and inherent significance of the voyage. The Trend tool graph of **way** (90), **turned** (80), **started** (118), **long** (107), **leave\*** (103) and **came** (122) is shown in Figure No. 29.

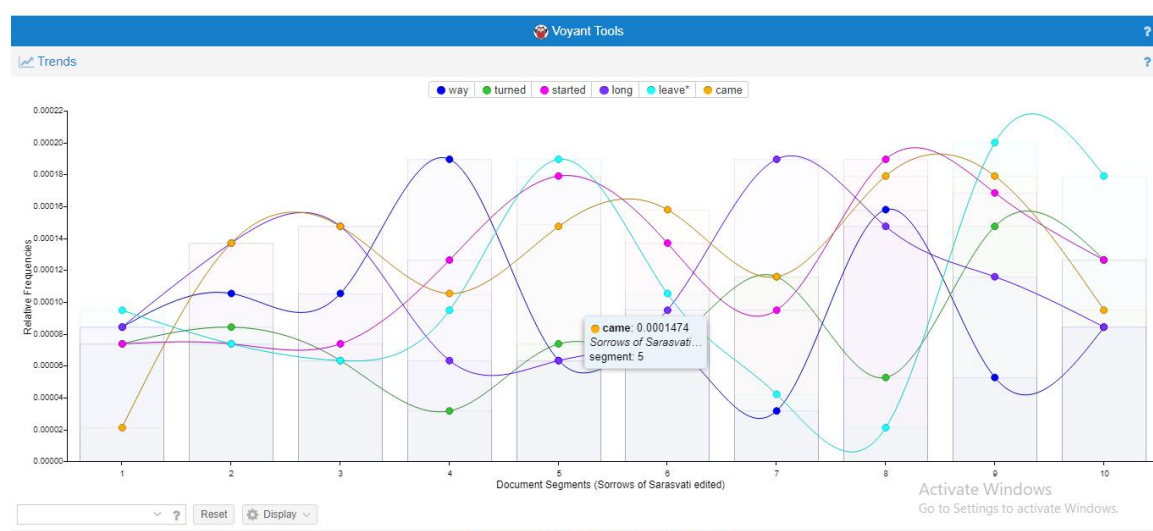


Figure 29. Trends Tool Graph of Theme of Journey

### 4.13 Theme of Human Fragility

**Like** (352), the third most frequent word in the corpus is sometimes used to represent simile. The term **like** (352) also contributes significantly to the theme of human fragility in nature. The use of the word **like** (352) contributes to the theme of human fragility by establishing a simile that emphasizes the intense and desperate condition of the individual. The phrase, “**She was panting like (352) a thirsty hatchling**”, portrays the protagonist’s physical state, **like** (352) her exhaustion to that of a dehydrated baby bird (Figure No. 30). This comparison conveys a sense of vulnerability, weakness, and fragility, highlighting the immense strain and hardship the character is enduring. Additionally, the use of **like** (352) to describe the **sun** (73) as “**penetrating**” suggests that its intensity is so strong that it metaphorically pierces through the surroundings.

Voyant Tools			
Contexts			
Document	Left	Term	Right
<input type="checkbox"/>	Sorrows...	only a few drops were left to keep his heart beating. His tongue was stuck,	like a dead snake, against his palate. Gasping and panting like a thirsty hatchling, he was he flapped them under the scorching sun. The red moisture in his veins was burned to a mere dryness; perhaps only a few drops were left to keep his heart beating. His tongue was stuck, like a dead snake, against his palate. Gasping and panting like a thirsty hatchling, he was struggling to keep the dead weight fluttering. Still there was some energy left in his eyes, flickering like last flames
<input type="checkbox"/>	Sorrows...	beating. His tongue was stuck, like a dead snake, against his palate. Gasping and panting	like a thirsty hatchling, he was struggling to keep the dead weight fluttering. Still there was in his veins was burned to a mere dryness; perhaps only a few drops were left to keep his heart beating. His tongue was stuck, like a dead snake, against his palate. Gasping and panting like a thirsty hatchling, he was struggling to keep the dead weight fluttering. Still there was some energy left in his eyes, flickering like last flames. He tried to fold his claws against his belly to

Figure 30. The theme of Human Fragility through Like

**Pakli** (98) and **Kiln** (161) elaborate on the fragility of human nature, particularly in “Beyond **Pakli’s** (54) **kiln** (161), there was a **field\*** (151) for burial; the deceased was buried in clay coffins” (Figure No. 31). The observation reinforces the overarching theme of human fragility, as it underscores the delicate and vulnerable nature of the human life and characters. This imagery linked with **like** (352) further reinforces the theme of human fragility by illustrating the oppressive and overwhelming nature of the environment and its impact on the characters. Overall, the inclusion of **like** (352) creates vivid comparisons and imagery that emphasize human fragility, depicting the physical and emotional challenges faced by the characters in the narrative (Figure No. 30).

Voyant Tools			
Contexts			
Document	Left	Term	Right
<input type="checkbox"/>	Sorrows...	and Sukra, were occupied with their work. Pakli’s upturned vessels were neatly arrayed in the	kiln . The spaces between them were stuffed with potsherds and ashes. Her children had placed their have always been like this.” Parushni stomped angrily out and started walking towards the kiln.
<input type="checkbox"/>	Sorrows...	did not lift their heads to see her face. The smoke emitting from the burning	kiln was soaring higher. “Didn’t you bake idols of Maha Mayya?” Parushni asked. “No, Pakli makes with a few beads. Pindoo and Sukra were so occupied with their work that despite noticing Parushni’s moving feet, they did not lift their heads to see her face. The smoke emitting from the burning kiln was soaring higher. “Didn’t you bake idols of Maha Mayya?” Parushni asked. “No, Pakli makes and bakes them herself,” one of the boys answered without looking up. Parushni adjusted the ewer on her hips and
<input type="checkbox"/>	Sorrows...	to procure a large piece of rock which I found near the woods beyond Pakli’s	kiln . I have removed it from the rock, and after cutting it into this shape, I in delight, “I, Samroo, have cut and styled this gem. I was drenched in sweat and labored in sweltering heat to procure a large piece of rock which I found near the woods beyond Pakli’s kiln . I have removed it from the rock, and after cutting it into this shape. I sanded it, lapped it and polished it. I have heated it in the brackish water to turn it white. I
<input type="checkbox"/>	Sorrows...	the itinerant traders by bartering it with the food items. Outside the village, beyond Pakli’s	kiln , there was a field for burials; the deceased were buried in clay coffins. There was a task for common good. There was no furnace in the village to melt copper; they would acquire copper from the itinerant traders by bartering it with the food items. Outside the village, beyond Pakli’s kiln , there was a field for burials; the deceased were buried in clay coffins. There was a pebbled path alongside it. After leaving the clay vessel in the field, relatives of the dead would place a

Figure 31. Term Kiln shows various Themes

Using the Corpus Collocates tool, the word **like** (352) reveals that it predominantly collocates with the term **animals\*** (94). This collocation suggests that human beings are frequently likened to **animals\*** (94) rather than depicted as integrated and resilient individuals. This portrayal accentuates the theme of loss and sorrow by showcasing human fragility. As Awan and Khan (2021) explore the fragile nature of humans- the reliance on animal comparisons becomes apparent. The collocate frequency of **like** (352) with **animals\*** (94) is found to be 15. At the same time, its co-occurrence with the two novel protagonists, **Virchan** (364) and **Parushni** (311), amounts to 13 each. The Voyant Trends tool further elaborates on these three collocates of **like** (352), represented through a graph obtained from the Trends tool (Figure No. 32).

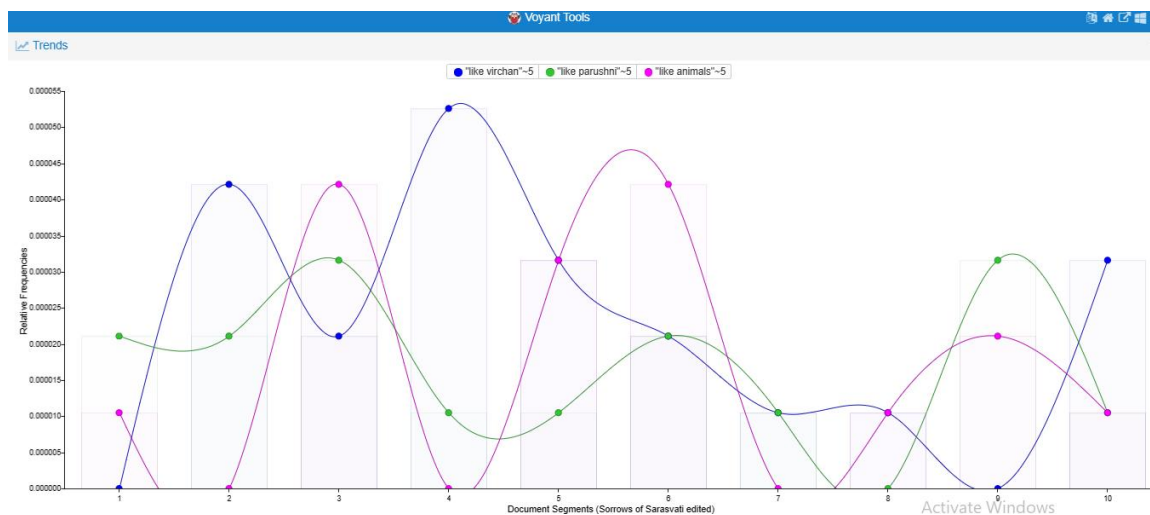


Figure 32. The theme of Human Fragility through a Graph of Collocates of Like

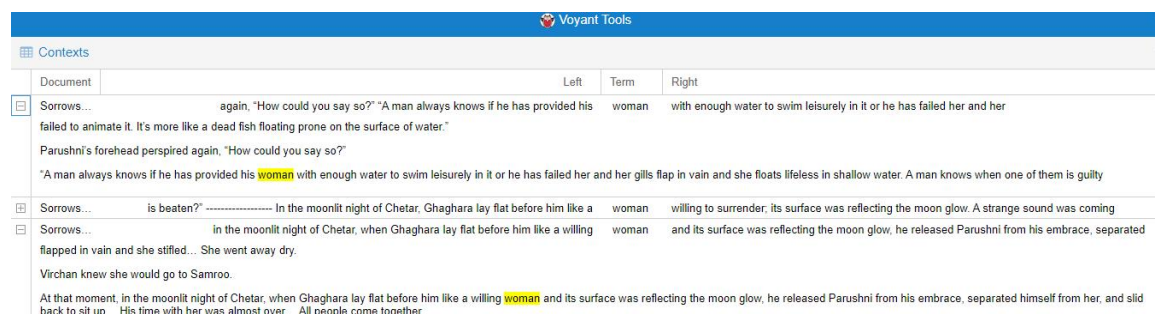
#### 4.14 Theme of Communication and Knowledge

The terms **said** (152), **say** (68), **saying** (15), **say** (9), **asked** (99), **ask** (34), **inquired\*** (7), **screamed\*** (50), **replied** (44), **reply** (7), **explained\*** (7), **just** (84), **yes** (84), **not** (543), **no** (235) and **none** (35) display the theme of communication because the story's characters talk to each other and about **people** (83). With dialogues, the plot of the story progresses, and all characters evolve and understand situations minutely. The theme of knowledge is supported by the words **know** (154), **saw** (127), **day** (118), **felt** (102), **knew** (93), **got** (83) and **looked** (78). These are fundamental and universal symbols for otherworldly divinity, spirituality, enlightenment and knowledge. The words are the wellspring of goodness and definitive reality. These are also the principal images of freedom, hope and goodness. Morality, purity, virtue, and values are associated terms too. Moreover, WH questions **When** (283), **Who** (249), **What** (223), **Where** (172), **Why** (160), **How** (148), **Which** (101), **Whose** (27) and **Whom** (16) also present the theme.

#### 4.15 Theme of Sex and Marriage

The theme of sex and marriage is one of the suggestive themes included in the novel. It is present within the text — directly and indirectly. Contexts tool revealed various instances of sexual encounters of human-to-human and human-with-animals. The theme of sex is present in the text, “A man always knows if he has provided his woman with enough **water** (458) to swim leisurely in it or he has failed her and her gills flap in vain and she floats lifeless in shallow **water** (458)” (Figure No. 33). Other instances are as follows, “You’re still stuck in Samroo’s thought; and it inhibits your liquids to flow when we make love. I know it”; “She no longer cared why the three of them —

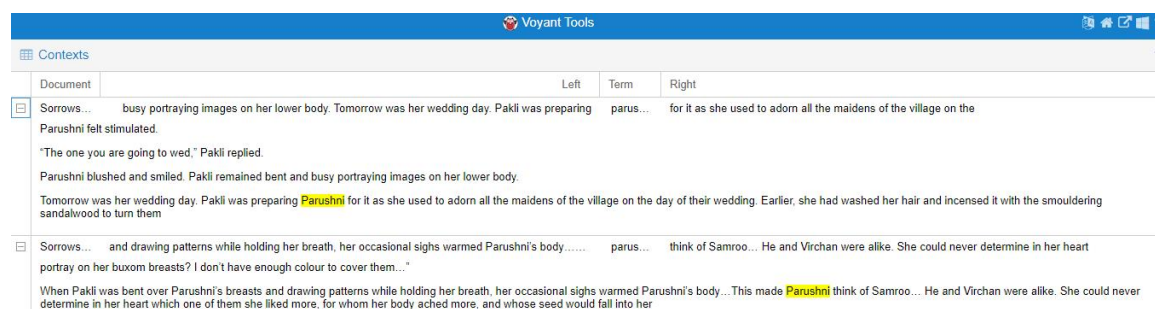
**Wango** (5), **Cheeva** (83), and **Jhoorya** (8)— took turns to be intimate with her during the **bull** (87) chase. ...Such willing surrender was a norm in the **village** (269)” and “He released **Parushni** (311) from his embrace, separated himself from her, and slid back to sit up... His time with her was almost over...” (Figure No. 33).



Document	Left	Term	Right
<input type="checkbox"/>	Sorrrows... again. "How could you say so?" "A man always knows if he has provided his failed to animate it. It's more like a dead fish floating prone on the surface of water." Parushni's forehead perspired again. "How could you say so?" "A man always knows if he has provided his <b>woman</b> with enough water to swim leisurely in it or he has failed her and her gills flap in vain and she floats lifeless in shallow water. A man knows when one of them is guilty	woman	with enough water to swim leisurely in it or he has failed her and her
<input type="checkbox"/>	Sorrrows... is beaten?" ----- In the moonlit night of Chetar, Ghaghara lay flat before him like a	woman	willing to surrender; its surface was reflecting the moon glow. A strange sound was coming
<input type="checkbox"/>	Sorrrows... in the moonlit night of Chetar, when Ghaghara lay flat before him like a willing flapped in vain and she stifled... She went away dry. Virchan knew she would go to Samroo. At that moment, in the moonlit night of Chetar, when Ghaghara lay flat before him like a willing <b>woman</b> and its surface was reflecting the moon glow, he released Parushni from his embrace, separated himself from her, and slid back to sit up... His time with her was almost over... All people come together	woman	and its surface was reflecting the moon glow, he released Parushni from his embrace, separated

Figure 33. Theme of Sex revealed through Contexts Tool

The theme of marriage is also evident in the text. Instances of the theme of marriage include, “Tomorrow was her wedding day. **Pakli** (98) was preparing **Parushni** (311) for it as she used to adorn all the maidens of the **village** (269) on the day of their wedding” (Figure No. 34), “**Samroo** (183) himself led his marriage procession, comprising all men of the **village** (269), and **Parushni** (311) sat waiting in the yard of her thatched hut surrounded by women with burning oil lamps” and “This made **Parushni** (311) think of **Samroo** (183)... He and **Samroo** (183) were alike. She could never determine in her heart which one of them she liked more, for whom her **body** (242) ached more, and whose seed would fall into her **like** (352) a stone goes and stays into a well” (Figure No. 34).



Document	Left	Term	Right
<input type="checkbox"/>	Sorrrows... busy portraying images on her lower body. Tomorrow was her wedding day. Pakli was preparing Parushni felt stimulated. "The one you are going to wed." Pakli replied. Parushni blushed and smiled. Pakli remained bent and busy portraying images on her lower body. Tomorrow was her wedding day. Pakli was preparing <b>Parushni</b> for it as she used to adorn all the maidens of the village on the day of their wedding. Earlier, she had washed her hair and incensed it with the smouldering sandalwood to turn them	parus...	for it as she used to adorn all the maidens of the village on the
<input type="checkbox"/>	Sorrrows... and drawing patterns while holding her breath, her occasional sighs warmed Parushni's body..... portray on her buxom breasts? I don't have enough colour to cover them..." When Pakli was bent over Parushni's breasts and drawing patterns while holding her breath, her occasional sighs warmed Parushni's body... This made <b>Parushni</b> think of Samroo... He and Virchan were alike. She could never determine in her heart which one of them she liked more, for whom her body ached more, and whose seed would fall into her	parus...	think of Samroo... He and Virchan were alike. She could never determine in her heart

Figure 34. Theme of Marriage through Parushni and Virchan

#### 4.16 Patriarchy versus Matriarchy

Many critics believe that the novel is Matriarchal in nature. It has a female protagonist **Parushni** (311) and the story revolves around her. However, there are instances in the novel that are patriarchal in nature. First, the female protagonist **Parushni** (311) is talked about less than the male character **Virchan** (364). Further, the male characters are more in number compared to female characters. This means that the male

characters are given more weightage compared to the female characters. The prominent male characters include **Virchan** (364), **Dorga** (252), **Samroo** (183) and **Mamam Masa** (137), while the prominent female characters include **Parushni** (311), **Pakli (Palki)** (98), **Gagri** (53) and **Mati** (22). **Man** (47) is also more in number compared to **woman** (21). Clearly, the male characters outnumber the female characters in the novel.

There are many instances of patriarchy in the novel. One of the instances includes, “**Mana** (9) has granted more power and wisdom to woman. And you know **Maha Mavva** (9), the supreme goddess, is also a woman” (Figure No. 35). In this example, it is evident that the people of the village believed that power and wisdom were granted by **Mana** (9), a male god in their mythology. Another instance includes, “**Males are reserved only for some trivial tasks or for breeding**” (Figure No. 35). Although males were given the tasks for breeding or some trivial things, women were entrusted with all the tasks as **Parushni** (311) used to bring water for the village and **Pakli** (98) used to work in the **kiln** (161). Women also brought about their children as **Mati** (22) had three sons all named **Jhoorya** (8) [or also called **Wango** (5), **Chandroo (Chandru)** (2) and **Jhoorya** (8)]. Men usually go to visit places and travel, like **Virchan** (364). Amidst claims of a matriarchal society, this indicates that the society was actually patriarchal in nature.

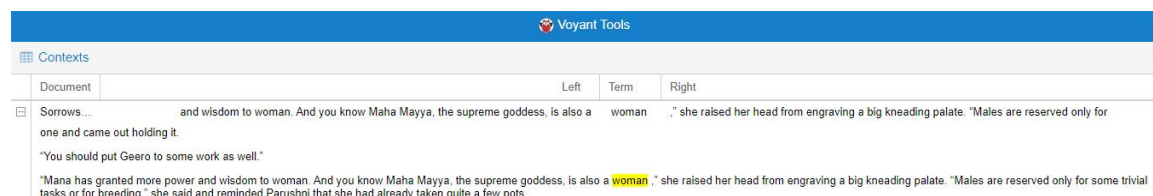


Figure 35. Instance for Patriarchy versus Matriarchy Debate

#### 4.17 Significance of Anatomy

In the novel, the translators employ a selection of words related to the anatomy of animals, birds and humans which include **body** (242), **eyes** (161), **face** (86), **head** (149), **hand** (84) and **beak** (8) to intricately convey the diverse gestures and expressions exhibited by these captivating avian creatures. The recurrent mention of the **body** (242) invites us to envision their overall posture and physical presence, providing insights into their graceful forms as they explore and navigate their environment with agility and elegance, as the writer describes parching body, inflamed body and buxom body. Furthermore, the deliberate emphasis on the **eyes** (161) holds significant relevance as it draws attention to the birds' acute visual perception. Birds often possess exceptional eyesight, and through the depiction of their eyes, such as “**black eyes** (161)”, “**parched**

**eyes** (161)” and “**shrinking eyes** (161)”, the writer conveys their ability to keenly observe their surroundings and communicate through subtle yet meaningful eye movements. Through these expressive eyes, the birds and characters truly connect with their environment, capturing the reader’s imagination.

The inclusion of **face** (86) delves into the inner world of the avian creatures, allowing us to contemplate the features and expressions displayed on their faces. By considering their beaks, eyes, and feathers, one gains insight into the birds’ emotions and the various ways they communicate. The face becomes a canvas upon which the birds express their unique personalities and inner feelings, as narrated, “... **face also turned as mysterious as a marsh, hiding dangers beneath the apparent calm**”. The mention of the **head** (149) provides a lens through which one explores the cognitive abilities of birds. Their brains, eyes, and ears lie within their heads, forming a vital centre for information processing, threat assessment, and quick decision-making. By emphasizing the significance of the head, the translators illuminate the intelligence and strategic thinking of these avian creatures, as mentioned in the novel, “**shook his head** (149) **sideways.**”

Lastly, the inclusion of the word **hand** (84) adds an intriguing dimension to the theme of bird anatomy (Figure No. 36). While birds do not possess hands in the traditional sense, their wings and talons serve as their equivalent. Birds’ wings exhibit remarkable versatility, allowing them to soar gracefully through the skies, navigate intricate aerial manoeuvres, and captivate us with their awe-inspiring displays. These carefully chosen terms paint a vivid and awe-inspiring portrayal of the gestures and expressions of birds, inviting readers to immerse themselves in the captivating world of avian creatures. Through the repeated usage of these terms, a theme centred around bird anatomy emerges, allowing readers to vividly imagine the intricate movements and behaviours of the birds being portrayed as major and minor characters are birds. By delving into their physical characteristics, sensory abilities, and unique adaptations, a deeper understanding and appreciation for the intricate marvels of bird anatomy is gained, leaving readers with a profound sense of wonder and admiration.

Voyant Tools			
Contexts			
Document	Left	Term	Right
Sorrows...	said, "You too have come to the lake to die?" The bird wilted in her	hand	. He was dead. "Meow, Meow," a peacock screamed from somewhere in the woods. She placed
	heap of dead birds in her one hand and, while dipping water into his open beak with the other, she said, "You too have come to the lake to die?"		
	The bird wilted in her	hand	. He was dead.
	"Meow, Meow," a peacock screamed from somewhere in the woods.		
	She placed that handful of lifeless feathers on the decaying pile of dead birds and stood up.		
	"Big waters used to flow		

*Figure 36. Significance of Anatomy*



## 4.18 Interplay Between Characters and the Natural Environment

The interplay between human existence and the natural environment also emerge from the most frequent words like **water** (458), **river** (339), **ghaghara** (162), **bank** (139), **sand** (124), **clay** (100), **soil** (89), **surface** (80), and **ground** (78). This theme highlights the intrinsic connection between survival, culture, spirituality, and the ever-changing dynamics of the ecosystem. **Water** (458) symbolizes both the physical substance of **life** (69) and the spiritual realm, embodying renewal, fear, comfort, and danger. The **river** (339), particularly the **ghaghara** (162), acts as a geographical and cultural anchor, shaping the landscape and influencing the daily lives of the characters. It represents movement, change, division, and the boundaries that define identities. Consequently, the combined frequencies of **Ghaghara** (162) and **river** (339) highlight their significance in conveying the theme of nature and its untamed essence (Figure No. 37).

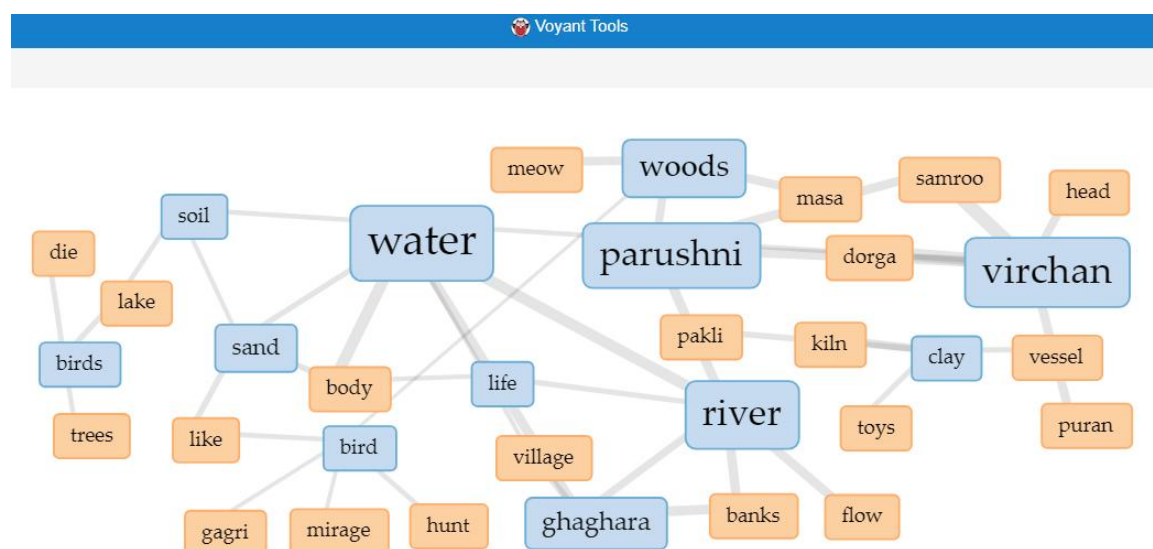


Figure 37. The interplay between Characters and the Natural Environment

Moreover, **water** (458) frequently co-occurs with the term **river** (339) to symbolically convey the essence of nature and its untamed characteristics. This connection between **water** (458) and **river** (339) exemplifies the prevalence of this thematic association. The collocate of the **river** (339) appears close to **water** (458). It has been observed a total of 19 times, further highlighting the significance of this relationship. Figure No. 38 illustrates the frequency trends associated with the simultaneous presence of **water** (458) and **river** (339). The comparison between **Virchan** (364) and **Samroo** (183) exhibits their connection between different characters. The parallelism expressed through “and **Samroo** (183) too...he was **like** (352) **Virchan** (364) for her... and **Virchan** (364) was **like** (352) **Samroo** (183)...” suggests that both characters were equal for Parushni.

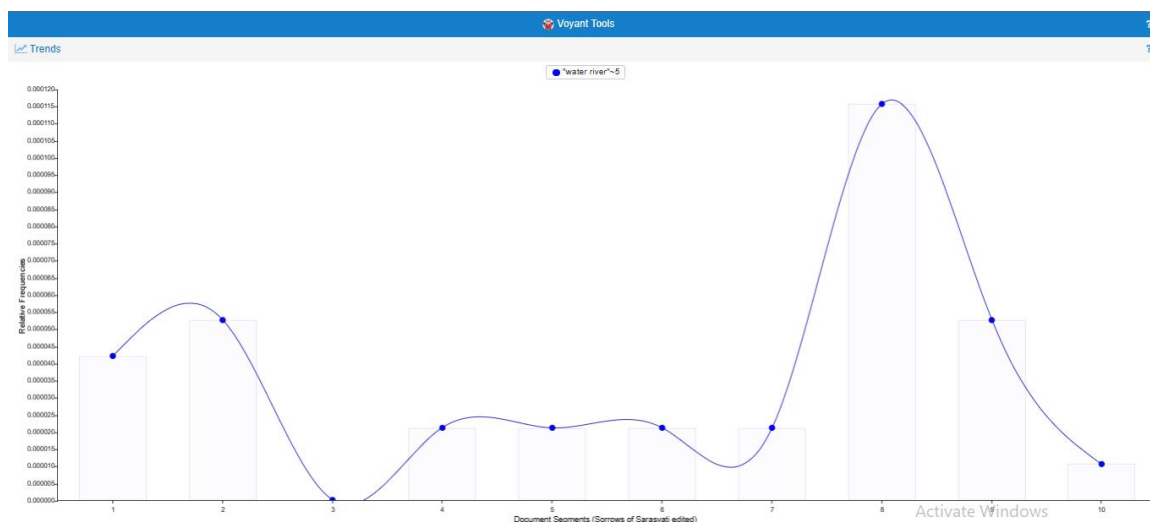


Figure 38. Graph of Collocation of Water and River through Trends Tool

Within the corpus, **Virchan** (364) and **Samroo** (183) emerge among the most frequently co-occurring words throughout the novel. As depicted in the corpus, the collocation frequency of **Virchan** (364) with **Samroo** (183) is recorded as 39, indicating that **Virchan** (364) is predominantly associated with **Samroo** (183) in the novel (Figure No. 4). The repeated occurrence of **Virchan** (364) and **Samroo** (183) together suggests a significant connection between these two entities within the narrative. The high collocation frequency implies a strong thematic association between **Virchan** (364) and **Samroo** (183). The Trends tool presents the co-occurrence patterns between **Virchan** (364) and **Samroo** (183). A notable observation is the overall upward trend in segments 5, 8, and 9, indicating the highest occurrence of co-occurrence in these particular segments. Similarly, the concordance of **water** (458) and **body** (242), as well as the concordance of **Parushni** (311) and **Virchan** (364), is also significantly high in segments 5, 8 and 9. The other segments exhibit a lower frequency of co-occurrence (Figure No. 39).

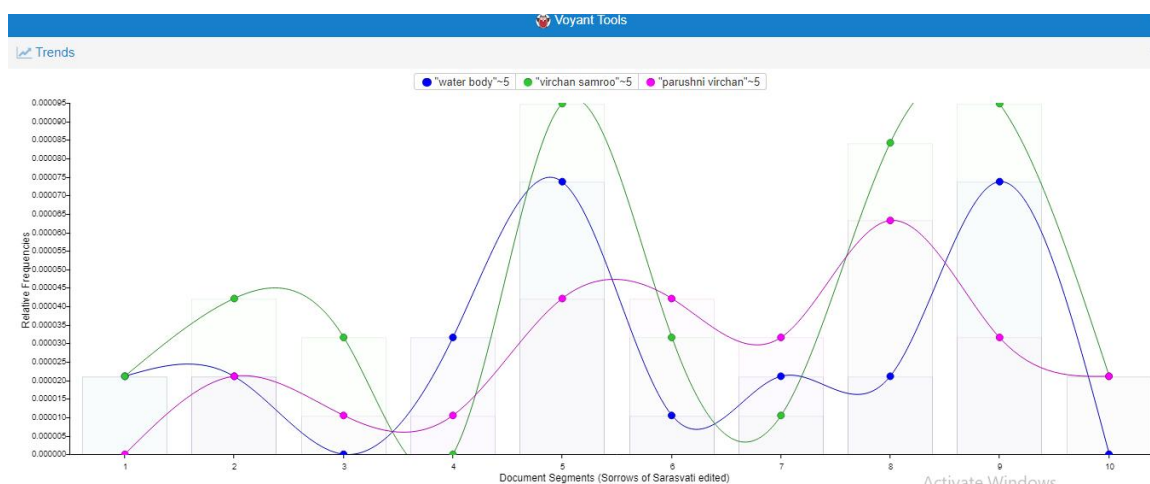


Figure 39. Significant Trends in Collocation patterns of major characters

## CHAPTER 5

### DISCUSSION

This chapter provides a discussion of findings and their relation to thematic analysis. It also discusses the main results of the study and the answer to research questions posed in Chapter No. 1. Through TM of CSSLR, the present research was designed to achieve five (5) research objectives and answer their associated research questions, which were specific and measurable and were focused on the outcomes of the study. The inquiries in the form of research objectives and research questions initiated in the first chapter are answered in this chapter along with establishing the link of research objectives and research questions with the findings and thematic analysis.

#### **5.1 Summary Tool: Frequency Distribution of Most Frequent Words**

The first research question was related to the use of the most frequent terms in the CSSLR obtained through the Summary tool. The findings revolved around the implementation of the summary tool to extract the most frequent terms from the novel *Sorrows of Sarasvati: The Lost River*. The findings demonstrate that this tool successfully identified fifty-nine (59) most commonly used words. The utilization of the summary tool helped with valuable insights into the prominent themes and content presented within the CSSLR. For instance, the term water stands out as the most prevalent, appearing a total of 458 times throughout the novel. The analysis revealed that **water** (458) was significantly linked to themes of death, nature, wilderness, loss, sorrow, mystery, and the unknown. Thus, the summary tool presented a hierarchical pattern of frequently used terms along with their frequencies, facilitating the prediction of themes. Thus, the Summary tool effectively addressed the first research question and fulfilled the first research objective by providing a comprehensive list of the most frequent words.

#### **5.2 Cirrus Tool: Word Cloud Visuals and Theme Insights**

The second research objective and its related research question were related to the portrayal of the significant terms from *Sorrows of Sarasvati: The Lost River* based on their statistical weight in the form of word cloud visuals. The word cloud visual of the most frequent terms is displayed in Figure No. 02. The visual of these terms was based on

their statistical weight. The terms **water** (458), **Virchan** (364), **like** (352), **river** (339), **Parushni** (311), **village** (269), **Dorga** (252), **body** (242), **come** (238), **woods** (221), **Samroo** (183), and **Ghaghara** (162) are more prominent in the picture because of their frequent occurrence in the novel. The less frequent words are shown with less emphasis because these terms obtained through TM have less significance in the novel and portrayal of themes compared to the most frequent terms. These visuals are also significant because they cast an image in the minds of readers, which is easier to remember. Therefore, the Cirrus tool is useful for portraying word cloud visuals based on their statistical weight, which portrays the themes of the novel.

### 5.3 Collocates Tool: Significant Collocations and Linguistic Patterns

The third research objective and its related research question were linked to the Corpus Collocates tool. Collocations are word combinations that frequently occur together within a text. Collocations play a crucial role in understanding the themes of a corpus by revealing the contextual relationships between words and shedding light on the semantic nuances within a text. These word pairs or groups often share a close semantic relationship or have a particular connotation when used together. By identifying and analyzing collocations, the contextual significance of specific terms within the corpus is revealed. This tool was helpful in the portrayal of themes from the novel with the help of collocations, for instance, the collocates of **water** (458) and **Virchan** (364) portrayed the theme of mystery and unknown (Figure No. 4) and collocates of **dead** (78) portrayed the theme of death (Figure No. 15). Collocates of **like** (352) were also significant for portraying various themes in the corpus. The Corpus Collocate tool has depicted collocation patterns of themes in the novel *Sorrows of Sarasvati: The Lost River*, which were significant for the portrayal of significant terms and their associations.

Collocations also serve as thematic markers within a corpus. When certain words consistently collocate with each other, they may indicate recurring themes or motifs in the text. Collocations also carry subtle nuances and connotations that may not be immediately apparent when examining words in isolation. They provide insight into how specific terms are used within the context of the text, offering a more nuanced understanding of the themes. Collocations are also used to analyze characters, their relationships, and narrative dynamics. When analyzing collocations related to specific characters or narrative elements, researchers can gain insights into character traits, plot developments, and thematic motifs. This allows for a more in-depth exploration of how characters and

narratives contribute to the overall themes of the text. Collocations serve as valuable linguistic tools for uncovering and understanding the themes of a corpus. They provide context, highlight recurring patterns, and reveal the semantic intricacies of a text, ultimately enhancing the researcher's ability to dissect and interpret the underlying themes within a literary work (Szudarski, 2023).

#### **5.4 Trends Tool: Trends in Word Usage**

The fourth research objective and research question were related to the depiction of key themes in CSSLR through the Trends tool graphs, distributed within ten (10) segments. In the Voyant suite, the themes are not only depicted through graphs of words but also the graphs of collocations were also significant for the portrayal of knowledge patterns. Such knowledge empowers readers with an understanding of the prevailing themes and predictions regarding their presence in the novel. Consequently, terms with higher frequencies in a segment were anticipated to exhibit stronger associations with certain themes compared to other segments in the corpus.

#### **5.5 Contexts Tool: Contextual Analysis of Terms**

There were different terms in the corpus which were explored with the help of their contexts, particularly the nomenclature. Various frequent words were used in various contexts in the novel, and these words portrayed various themes. This tool was very helpful for tracking the connection between words and their pragmatic meanings. The fifth objective was fulfilled and the subsequent fifth research question of this research was fulfilled as the contexts tool portrayed themes within the novel. Contexts tool revealed nomenclature and provided basics to qualitative thematic analysis.

#### **5.6 Integration of Findings with Research Objectives and Questions**

The themes mentioned in the text *Sorrows of Sarasvati: The Lost River* are closely connected to the research objectives and research questions through the process of TM and thematic analysis. The themes extracted from the text are connected to these objectives. The first question is directly linked to the themes as it seeks to understand the frequency distribution of thematic words in the text. The themes identify the most frequent words related to these themes, the research aims to reveal the frequency distribution of thematic elements within the text. The frequency distribution of the most frequent words through the Summary tool reveals the Proto-Dravidian terms in the corpus.

The research objective is to reveal the frequency of the most common words in the text using the Summary tool.

The second research question focused on using word clouds to visually represent themes, making it directly related to the identified themes in the text. Word clouds generated using the Cirrus tool visually represent the prominence of words in the text. The themes, like the theme of nature and wilderness or emotions, sensations, and nostalgia, are depicted through word cloud visuals, providing insights into their significance in the narrative. The objective of the Cirrus/ word cloud is to create word cloud visuals that provide insights into themes. The unique character names contribute to the distinctive themes in the narrative. Visual representations of these names in word clouds emphasize their significance.

The third question dealt with the extraction of significant collocations that can uncover how thematic words are associated with other words, aligning with the thematic analysis. The thematic analysis, including themes like the interplay between existence and the natural environment or the theme of the journey, can be further explored through the extraction of collocations. Collocations revealed how specific words and themes co-occur, shedding light on their interconnectedness in the text. The collocations extracted the connections of key terms and their significance in data analysis. Collocation extraction through the Corpus Collocation Tool is important in the generation of results. The objective of the Corpus Collocates tool is to extract collocations (word pairs). The unique character names, when collocated with other terms, reveal specific linguistic patterns that are unique to the text and Proto-Dravidian language.

The fourth question tried to find the word usage in the graphical representation to unveil the trends in the data analysis using the Trends Tool. The identification of trends in word usage, which included thematic words, allows for an understanding of how themes evolve throughout the text. The themes of impermanence and transience or community and shared responsibility may be manifested as trends in the text. Trends Tool graphs helped in identifying significant trends in the corpus. The objective of the Trends tool was to identify trends among the most frequent words. The repeated appearances of unique character names indicated trends in the narrative related to these characters.

The fifth question directly explores the relationship between words and themes in different contexts, supporting the contextual analysis of themes. The research objectives

related to understanding the contexts of different words and themes for the portrayal of themes directly align with the themes of communication and knowledge or the theme of human fragility. Analyzing word contexts provides insights into how these themes are woven into the narrative. Contexts tool helped in understanding word contexts. The objective is to understand how different words function in various contexts to portray themes. Character names represent not only individuals but also the contexts and themes associated with them.

### **5.7 Novelty in Findings**

The presence of unique and frequently occurring terms highlights the importance of novel terms in portraying the themes. These names represent individual characters and contribute to the uniqueness of the narrative. The nomenclature in the text is unique because it encompasses a range of distinct elements, including character names, place names, deity names, and even names of seasons, months, plants, and garments. These names serve as key identifiers within the narrative, contributing to the text's individuality. The male characters possess unique personalities and traits, making them stand out in the text. The repeated use of these names establishes a sense of familiarity and importance, making them idiosyncratic to the story. Although appearing less frequently, the female character names also add depth and balance to the cast, contributing to the overall richness of the narrative.

The names of the mentioned places serve as important settings in the story. The repeated references to these locations create a sense of familiarity and connect readers to the unique world depicted in the text. The names of rivers are not only rivers but also significant elements in the story's thematic and cultural context. These names represent more than just geographical features; they hold cultural and historical significance. The inclusion of names related to seasons, months, plants, and garments also adds cultural depth to the text. These terms are not just labels but contribute to the overall portrayal of the setting and atmosphere. The nomenclature in *Sorrows of Sarasvati: The Lost River* is unique and idiosyncratic because it goes beyond mere labels; these names hold cultural, thematic, and narrative significance, contributing to the distinctive character of the text.

Foregrounding the question of thematic analysis entails bringing the central inquiry regarding thematic exploration to the forefront of the research narrative. It entails stating the primary questions that direct the examination of the thematic components

clearly and concisely. The research creates a precise and targeted framework for investigation by emphasizing the theme analysis issue, giving readers a road map for comprehending the main goals and objectives of the investigation. This foregrounding serves to orient the reader towards the core thematic concerns of the research, setting the stage for a systematic exploration of thematic motifs, narrative themes, and conceptual frameworks embedded within the text.

## 5.8 Significance of Stylometry

Stylometry, the study of an author's writing style, can reveal valuable insights about the author's choices and tendencies in writing. In the context of *Sorrows of Sarasvati: The Lost River* and considering the research questions and objectives, the exploration of the nomenclature, themes, and textual patterns revealed the translators' quantitative stylometry. Frequency distribution of most frequent words through the Summary tool focused on the frequency of the most frequent words, and it was also applied to character names and unique terms. Stylometry involves identifying consistent patterns in word usage, which extends to naming conventions. The themes in *Sorrows of Sarasvati: The Lost River* are novel due to their cultural, linguistic, and emotional depth, as well as their distinctive portrayal within the narrative.

The translators' stylometry is also an important factor to consider in this research. Translators unintentionally or intentionally bring their own linguistic choices and interpretations to the text. Their translation decisions, including how they convey and interpret the unique themes and nomenclature, can influence the overall stylometry of the translated work. The nomenclature, themes, and textual patterns within *Sorrows of Sarasvati: The Lost River* offered insights into the translators' stylometry. This includes their stylistic choices in character naming, thematic consistency, and linguistic patterns. Stylometry involves identifying and analyzing these consistent patterns to gain a deeper understanding of the translators' unique writing styles and preferences. The current study reveals the translators' stylometry with the help of TM and themes.

The creation of word clouds highlighted the translators' stylistic choices in emphasizing certain terms or themes. If character names or thematic terms appear prominently, it reflects the translators' stylistic emphasis on these elements. Stylometry often involves analyzing word pairs and their co-occurrence patterns. The choice of character names and how they collocate with other words can offer insights into the



translators' linguistic patterns. Identifying trends among the most frequent words uncovered the translators' consistent stylistic choices and preferences in terms of vocabulary. Analyzing how different words function in various contexts revealed the translators' attention to detail and nuances in language use, contributing to the understanding of their stylometry.

The consistent use of unique character names suggests that the translators have a penchant for distinctive character naming conventions. The names of all the things reflect the culture of the Proto-Dravidian era. This attention to character nomenclature reflects a stylistic choice in character development. Analyzing the collocations of character names with other words and how these collocations contribute to the narrative can shed light on the translators' narrative and stylistic choices. Stylometry seeks to identify these consistent patterns in translators' writing. The inclusion of deity names, place names, and terms related to seasons, months, plants, and garments showcases the translators' commitment to creating a richly detailed and immersive world within the text. This suggests a stylistic inclination toward world-building and cultural depth.

The translators' choice of themes, which can be inferred from the nomenclature (e.g., names related to rivers, places, and deities), reflects a thematic consistency in their writing. Stylometry often involves identifying recurring themes or motifs in translators' body of work. Themes related to nature (e.g., rivers, seasons, plants), culture (e.g., deity names, place names), and character relationships (e.g., unique character names) are indicative of the translators' thematic preferences and their ability to weave these themes into the narrative. The frequency of specific words, including character names, in relation to other words in the text reveals the translators' propensity for certain vocabulary and linguistic patterns, contributing to the understanding of their stylometry.

The themes from the novel are considered significant for several reasons. The themes are deeply rooted in the cultural and linguistic elements of the Proto-Dravidian context. This cultural specificity makes the themes unique and distinct from those found in other literary works. Themes related to nature, wilderness, transformations in the environment, and the interplay between existence and the natural world reflect a specific perspective on the environment and its significance within the narrative. Themes of sorrow, loss, nostalgia, and human fragility are explored in ways that may be distinct from conventional literary treatments. The emotional depth and nuanced portrayal of these themes contribute to their uniqueness.

The theme of mystery and the unknown, along with the exploration of death and impermanence, adds an enigmatic and distinctive dimension to the narrative. These themes may be approached in unique ways. Themes related to community, shared responsibility, and communication are idiosyncratic in the context of the story's setting and cultural backdrop. They reflect specific values and dynamics within the narrative. The inclusion of birds' anatomy as a theme is particularly distinctive. The themes represent the unique nomenclature and novel cultural beliefs within the narrative.

## **5.9 Relevance of Research Methodology with Findings**

The findings of this research, conducted within the framework of the research onion provide valuable insights into the text *Sorrows of Sarasvati: The Lost River*. These findings are presented through a combination of quantitative and qualitative methods, facilitated by various text mining tools. The use of post-positivist research philosophy signifies a commitment to empirical analysis while acknowledging the limitations of absolute objectivity. The quantitative results obtained through text mining align with the empirical orientation of this philosophy, providing concrete data on term frequencies, collocations, and trends. The findings contribute to a nuanced understanding of the text's content and linguistic patterns, in line with the empirical aspects of post-positivism.

The interactive research paradigm emphasizes the involvement of the researcher in the research process. In this study, the researcher actively engaged with TM tools to extract both quantitative and qualitative insights. The findings reflect the researcher's active role in data collection, analysis, and interpretation, demonstrating the application of the interactive paradigm. The abductive research approach combines deductive and inductive reasoning, allowing for the generation of hypotheses and theories from the data. In this context, the findings, particularly the themes identified through qualitative analysis, exemplify the abductive approach. The themes emerge from the data, contributing to a deeper understanding of the text's underlying narratives and literary elements.

The use of mixed methods, including quantitative and qualitative analyses, offers a comprehensive view of the text. The quantitative results, such as term frequencies and collocations, complement the qualitative findings, enhancing the richness of the research outcomes. This approach aligns with the holistic nature of mixed-method analysis, providing a multifaceted exploration of the text. The descriptive research strategy aims to provide an accurate portrayal of the phenomenon under study. In this study, the findings,

particularly the visualization of data through tools like Cirrus, Summary, Collocates, and Trends, serve the descriptive research strategy by presenting a clear and detailed account of the text's linguistic characteristics and thematic content.

The combination of quantitative and qualitative data in the findings reinforces the research onion's layered approach. The quantitative data, generated by text mining tools, offers statistical insights into word frequencies and associations, aligning with the outer layers of the research onion. Meanwhile, the qualitative data, including thematic analysis and word contexts, delves into the deeper layers of interpretation, contributing to a holistic understanding of the text's meaning and themes. The choice of a cross-sectional time horizon is aligned with the temporal layer of the research onion. This time horizon allows for the examination of data collected at a specific point in time, providing a snapshot of the text's content and linguistic patterns at that moment. The research findings are reflective of this temporal focus, offering insights into the text's characteristics without the need to track changes over time.

The descriptive research strategy, which aims to provide an accurate portrayal of the phenomenon under study, complements the research onion's methodological layer. In this case, the findings are descriptive in nature, as they present a clear and detailed account of the text's linguistic characteristics, thematic content, and other relevant aspects. This aligns with the research methodology's focus on describing and documenting the features of the text. The research findings, characterized by research onion, resonate with the specific layers of the research onion related methodology. These findings offer a focused and detailed analysis of the text *Sorrows of Sarasvati: The Lost River*, contributing to a comprehensive understanding of its content and characteristics within the chosen research framework.

### **5.10 Contribution of Theoretical Framework in Findings**

The findings of this research are closely aligned with the theoretical framework that blends Operations Management Theory (OMT) with digital TM and thematic analysis. The OMT, as introduced by Taylor in 2016, offers a theoretical lens through which to examine digitized learning. The relevance and contribution of the research findings to the theoretical framework is evident in the methodology applied. The research findings provide empirical evidence that supports and enriches the theoretical framework of OMT applied to digitized learning.

The research findings, which encompass both quantitative and qualitative analyses, directly correspond to the theoretical underpinning of OMT. OMT emphasizes the efficient and effective management of operations, and in this context, the quantitative analysis of digital TM exemplifies the operational aspect by leveraging computing power to extract meaningful insights from the text. The findings validate the application of OMT in the context of digitized learning by demonstrating how digital tools can enhance human interpretation and operational efficiency. The integration of quantitative and qualitative analyses aligns with OMT's principles of optimizing operations, and the findings demonstrate how digital TM and thematic analysis contribute to a deeper comprehension of the digitized learning context.

Quantitative analysis aligns with the digital TM of the research findings obtained through VT. The quantitative analysis conducted in the research, particularly the use of digital TM tools, aligns with OMT's focus on operational efficiency. By quantifying term frequencies, collocations, and trends within the text, the findings substantiate the idea that digitized resources, combined with linguistically sophisticated algorithms, can augment human interpretation. This resonates with the operational principles of OMT, which aim to optimize processes and resources.

Qualitative study is connected with the thematic analysis of the most significant terms, their trends and collocations obtained through VT. The qualitative component of the research, involving thematic analysis, connects with OMT's broader perspective. The thematic analysis serves as a framework for categorizing and understanding textual content based on recurring themes. In the context of OMT, this qualitative approach can be seen as a way to classify and organize thematic ideas related to digitized learning. The findings from thematic analysis enrich the theoretical framework by providing qualitative depth and insights into the themes emerging from the text.

The synergy between quantitative analysis (digital TM) and qualitative analysis (thematic analysis) exemplifies the holistic nature of the research. OMT encourages an integrated approach to operations management, and in this case, the integration of quantitative and qualitative methods mirrors the theoretical framework's emphasis on combining frameworks for digitized learning. The findings validate the relevance of this integrated approach by offering a comprehensive understanding of the digitized text. Thus, the use of a theoretical framework supports the generation of results and their explanation.

## CHAPTER 6

### CONCLUSION

#### 6.1 Reviewing the Current Procedures

This chapter provides a synopsis of the results, goals, and various contributions to the body of knowledge. The procedures and tools which helped in conducting the study are reviewed. Significant terminologies, frequencies and themes revealed through TM of VT are also précised in this chapter. The contribution of research in the field of study and suggestions for the future are also included in this chapter. VT is a free, publicly available web-based application for investigating novel forms of knowledge. Five (5) tools from twenty-nine (29) VT were merged in the current study: the Summary tool, the Cirrus/ word cloud, the Corpus Collocates tool, the Trends tool, and the Contexts tool. The statistical data of the fifty-nine (59) most common themes with their frequencies, a PNG image of a Cirrus/ word cloud with sixty-five (65) words, concordance, collocations, and trends of these frequent keywords were all included in the quantitative data. Whereas the analysis of the themes derived from the fifty-nine (59) notable words from the Summary tool, the description of the Cirrus/ word cloud visualizations, and their contexts made up the qualitative data.

The theoretical framework is the guiding structure for the study design. The theoretical framework for digitized learning was an operational model with two (2) frameworks. Taylor's (2016) Operational Management Theory (OMT) follows a digital TM approach (Ignatow, 2015) for the quantitative analysis, while the qualitative study includes a thematic analysis approach (Gibbs, 2012). Postulates of digital TM and thematic analysis were applied to find unfamiliar knowledge patterns and learning designs through the evaluation and interpretation of visual data and the validity of the findings. It is the extraction of implicit, previously unknown, and potentially useful information from data called a semantic thematic approach.

The major and important concepts within the novel were highlighted by Cirrus, which provides statistical evidence for these claims. In the Cirrus, the most common words and topics stand out in bigger fonts, while other terms are stressed less with small fonts. Moreover, its interactive nature provides statistical proof and permits individual

choice of 25-500 topics. Cirrus transforms static texts and corpora into a visually appealing and interactive resource that inspires curiosity and autonomy among the readers (Hayes, 2008). Scholarly applications of word clouds include the study of latent themes in areas, as diverse as catchphrase extraction, content investigation, nostalgic yearnings, correlation, forecast, jargon creation, vocabulary expansion, instruction, discourse analysis, synopsis output, the tendency of a creator, and inclination analysis.

The Summary tool provides an accurate visual depiction of the overall quantity and frequency of keywords across all corpora and a visual representation of frequent key themes throughout the full corpus. These characteristics indicate the translators' and the novel's computational stylistic characteristics or stylometry. The CSSLR has uncovered the stylometric traits in the novel through the application of the distant learning technique. Since Roberto Busa spent 22 years manually mining *Thomisticus Aquinas*, comparing this amount of time with the capability of a few seconds, the VT can extract stylometric components (Ullah, 2022). The Summary tool easily extracts the most frequent terms/ words and their frequencies from the corpus.

The Context tool helps readers understand any text better by pinpointing important words and providing detailed information about its sentence context, part of speech, and contextual meaning (Torre, 2019). This tool also simplifies information retrieval and extraction because of the methodical way it arranges critical data, making it possible for the reader to find key pieces of information within a large database quickly (McGowan, 2021). The Trends tool in VT is used to build the graphs of the most used terms. These graphs make it possible to see the most prevalent themes, significant changes, and core frequencies throughout the corpus (Neha, 2023). The terms, collocates and counts are displayed by the Corpus Collocate Tool. Terms are the keywords that need to be collocated, Collocates are the terms found surrounding each keyword, and Count is the frequency of the collocates (Sampsel, 2018).

Conclusively, the VT transformed raw data into meaningful and valuable knowledge patterns; for example, the Cirrus tool shows themes; the Summary tool exposes details about the corpus, unique words, and stylometric features; the Corpus Collocates Tool discloses the significant collocations; the Trends Tool portrays the trends of most frequent words; and the Contexts Tool reveals the contexts and concordance in the novel. These tools combined to obtain the qualitative data in the form of visuals, words, frequencies and collocates; and the quantitative data in the form of contexts of

terms and themes obtained through computational tools. Thereupon, this study ties the domains of CL, DH, and Pakistani literature in English.

## 6.2 Important Results

This key terms in the Corpus of *Sorrows of Sarasvati: The Lost River* obtained through TM are as follows: **water** (458), **Virchan** (364), **like** (352), **river** (339), **Parushni** (311), **village** (269), **Dorga** (252), **body** (242), **come** (238), **woods** (221), **Samroo** (183), **Ghaghara** (162), **eyes** (161), **kiln** (161), **know** (154), **said** (152), **fields** (151), **head** (149), **away** (139), **bank** (139), **time** (139), **Masa** (137), **Mohenjo** (129), **saw** (127), **sand** (124), **used** (123), **came** (122), **day** (118), **started** (118), **days** (111), **long** (107), **night** (107), **trees** (104), **felt** (102), **clay** (100), **went** (100), **asked** (99), **pakli** (98), **stood** (96), **tree** (95), **knew** (93), **left** (92), **way** (90), **soil** (89), **bull** (87), **face** (86), **years** (85), **hand** (84), **just** (84), **yes** (84), **Cheeva** (83), **got** (83), **people** (83), **surface** (80), **turned** (80), **sat** (79), **dead** (78), **ground** (78) and **looked** (78).

The statistical data of these terms helps to extract the nomenclature. Nomenclature includes the names, unique terminologies of the Proto-Dravidian language and cultural values that encourage deep linguistic analysis. Nomenclature is the compilation of the names of characters, places, god and goddess, rivers, lakes, seasons and months, plants and shreds and garments in the novel. Additionally, the following themes were explored from the TM of the CSSLR: the theme of nature and wilderness, the theme of mystery and unknown, the theme of superstitions, the theme of death, the theme of community and shared responsibility, the theme of loss and sorrow, the theme of emotions sensations and nostalgia, the theme of transformations in nature and environmental changes, the theme of impermanence and transience, the theme of journey, the theme of human fragility, the theme of communication and knowledge, significance of anatomy, and interplay between characters and natural environment (Danino, 2010).

The research objectives, grounded in the research onion and framed by the theoretical framework, provided a structured approach to exploring this complex novel. TM techniques were harnessed to extract quantitative data about word frequencies, collocations, and trends, and these findings were presented within the research onion's layered methodology. These findings, which included the quantitative results from TM and qualitative insights from thematic analysis, have a profound connection to the themes of the novel. The themes were unearthed through TM and substantiated through thematic

analysis. The research objectives and questions served as the guiding lights for this exploration, connecting the thematic essence of the novel with data-driven analysis.

The interplay between the research onion and the theoretical framework enriched the study by combining empirical and theoretical perspectives. Operations Management Theory (OMT) not only guided the research but also brought a fresh lens to understanding digitized learning. This connection underscores the relevance of theory to practice in the DH, bridging the gap between the novel's content and contemporary research methodologies. In essence, this research has demonstrated the fusion of literature, data analysis, and theory, emphasizing the interdisciplinary nature of DH research. This triangulation method is pivotal for comprehending *Sorrows of Sarasvati: The Lost River* and advancing research within the DH domain.

Valid information is gleaned from the Cirrus visuals and the statistics compiled using the Summary tool. The outcomes are consistent even after repeated iterations (McNaught and Lam, 2010). Voyant has the benefit of analysing frequency distributions rather than raw word counts to account for differences in document length. It was developed considering the DH, and user viewing and interpretative activities are meant to be made easier. VT produce word clouds from provided documents, displays word frequency, trends and collocation, and carries out additional TM operations. VT can be utilized to evaluate texts in various fields, notably literary works, language acquisition, advertisements and media industry, medical services, and information systems development (Anderson, 2023; Brown & Tor, 2023).

### **6.3 Contribution to the Field of Study**

Edison (2001) once said, “The value of an idea is in the use of it.” When a concept is implemented for the common good, it becomes useful and valuable to the people. Demonstrating the continuity between far and close reading, this successfully bridges the gap between the two. This investigation represents the first detailed analysis of *Sorrows of Sarasvati: The Lost River* using five (5) VT, including the Cirrus, Summary, Contexts, Trends, and Collocates tool. The novel is a compilation of stances and themes deeply woven into archaeology, sociology, culture, anthropology, history, and above all, linguistic features of the past, particularly the Proto-Dravidian period. This study contributes to the field of study in two (2) ways. First, the usage of VT for research and second, the extraction of themes of *Sorrows of Sarasvati: The Lost River*.



A few articles, some master's theses (Darwesh, 2020; Masood, 2022) and Ph.D. dissertations (Ullah, 2022) have used VT but not specifically on CSSLR. It exemplifies the potential of DH to be used in the pursuit of individual and contextualised knowledge acquisition and analysis. Text analytics, data visualisation, and information extraction are useful for students and teachers. This study would be useful for those studying English literature, linguistics and pedagogy. The students of English literature may now receive an overview of various works of Linguistics, Literature and Social Sciences before diving in, while the students of Linguistics will learn how to use VT in conjunction with TM. Students in Pakistan would be able to produce more digital material due to this research.

This work presents a novel method of literary analysis using corpus tools, which makes a significant addition to the realm of research. The study illustrates the potential of digital technologies in identifying thematic themes, linguistic patterns, and narrative structures within textual data by using computer approaches to evaluate a literary work. The incorporation of corpus tools represents a paradigm shift in the technique of literary analysis, opening up new avenues for the study of texts and enabling creative approaches to thematic analysis and textual interpretation. The study fosters interdisciplinary interaction and theoretical integration within the field of digital humanities research by contributing to its thematic and textual development through innovative methods.

In addition, putting the theme analysis issue front and centre helps place the importance and relevance of this study in the larger context of literary study and digital humanities conversation. It places the research into a broader intellectual framework by highlighting the theoretical framework and methodological technique that support the analysis of thematic components of the novel. The research highlights its dedication to revealing deeper levels of meaning and interpretation within the text by emphasizing the theme analysis inquiry. This enriches academic discourse and adds to the continuing discussion about thematic exploration in literary studies.

## **6.4 Future Research Directions**

Word clouds have a wide variety of potential applications in several fields, including ELT, lexicon education, vocabulary development, and the analysis of code-mixed language in multilingual corpora. In future studies, researchers may broaden their cloud-based word, emotion, and semantic searches in the same way. The generated word cloud might be utilised for the examination of many literary works and other disciplines,

or the findings could be used for real-time language instruction (Kilgarriff, 2014). Code mixing and code-switching might be analysed using Cirrus, Contexts, and Summary tools. The tools can also uncover linguistic patterns and lexical substitution, such as in Urdu, where many English words replace Urdu phrases. Moreover, many aspects of translation and fiction may be evaluated using text analytics tools like Cirrus, Summary, Contexts, Collocates, Trends, Phrases, Links, and TermsBerry.

To build upon the current research and address its limitations, several avenues for future exploration and inquiry are suggested. First, future research can undertake comparative TM analyses of multiple literary works or texts from diverse cultural backgrounds to identify common themes and unique variations. Second, integrating TM with other modes of analysis, such as sentiment analysis or critical discourse analysis, can provide a more comprehensive understanding of the text. Third, examining how themes evolve and are expressed in different cultural contexts can offer insights into the universality and cultural specificity of literary themes. Fourth, future research can leverage advances in ML and NLP to develop more sophisticated TM models for deeper thematic exploration. Fifth, exploring stylometry and authorship attribution techniques can help identify patterns unique to specific translators' within the broader corpus of literature. All these research directions can help other researchers in their work.

## 6.5 Research Implications

The findings and insights generated from this research carry several implications for the fields of literature, DH, and TM. First, the analysis contributes to a deeper understanding of the themes and linguistic patterns within *Sorrows of Sarasvati: The Lost River*, shedding light on the text, themes and their cultural context. Second, the research showcases the potential of DH methodologies, particularly TM, in uncovering novel themes and patterns in literary texts. Third, the integration of literary analysis with computational methods highlights the interdisciplinary nature of research and its capacity to bridge the humanities and technology. Fourth, the findings are valuable for educators and students interested in the study of literature and the application of computational tools for literary analysis. Fifth, the research contributes to the preservation and understanding of cultural and literary heritage by uncovering and documenting unique themes within the text of the novel.

## REFERENCES

- Adriaans, P., & Zantinge, D. (2009). *Data mining*. New Delhi: Pearson Education.
- Aggarwal, C. C., & Zhai, C. (2012). An introduction to text mining. In *Mining text data* (pp. 1-10). NY: Springer.
- Alhudithi, E. (2021). Review of Voyant Tools: See through your text. *Language Learning and Technology*, 25(3), 43-50.
- Alsulami, A. (2016). Review of challenges in big data mining. *American Research Journal of Computer Science and Information Technology*, 1, 1-4.  
<https://doi.org/10.21694/2572-2921.16005>
- Altameemi, Y., & Altamimi, M. (2023). Thematic analysis: A corpus-based method for understanding themes/Topics of a corpus through a classification process using long short-term memory (LSTM). *Applied Sciences*, 13(5), 3308.  
<https://doi.org/10.3390/app13053308>
- Anderson, A. (2023, February 30). *Using Voyant for text analysis*. Digital History Methods. Retrieved March 17, 2023, from <https://ricedh.github.io/02-voyant.html#:~:text=One%20advantage%20of%20Voyant%20is,for%20imbalance%20in%20document%20length>
- Archer, D., & Bousfield, D. (2010). ‘See better, Lear’? See Lear better! A corpus-based pragma-stylistic investigation of Shakespeare’s *King Lear*. *Language and Style*, 183-203. [https://doi.org/10.1007/978-1-137-06574-2\\_12](https://doi.org/10.1007/978-1-137-06574-2_12)
- Archer, D., Culpeper, J., & Rayson, P. (2009). Love – ‘a familiar or a devil’? An exploration of key domains in Shakespeare’s comedies and tragedies. *Taylor & Francis*, 157-178. <https://doi.org/10.4324/9781315547411-19>
- Argamon, S., Šarić, M., & Stein, S. S. (2003). Style mining of electronic messages for multiple authorship discrimination. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*.  
<https://doi.org/10.1145/956750.956805>
- Asif, M., Zaidi, S., & Yasmeen, F. (2021). A corpus-based study of Austen’s novel *Pride & Prejudice* using the Cirrus tool. *Corporum: Journal of Corpus Linguistics*, 4(1), 60-72. <https://journals.au.edu.pk/ojscre/index.php/crc/article/view/268>
- Awan, M. S., & Khan, S. (Trans.). (2021). *Sorrows of Sarasvati: The Lost River*. Pakistan Academy of Letters.

- Awan, M. S., & Khan, S. (Trans.). (2023). *Sorrows of Sarasvati: The Lost River*. Sang-e-Meel Publications, Lahore. ISBN No. 978-969-35-3517-4
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press ELT.
- Bhaskar, R. (2011). *Reclaiming reality: A critical introduction to contemporary philosophy*. Abingdon: Routledge.
- Biber, D., & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press.
- Blaikie, N. (2010). *Designing social research (2nd ed.)*. Cambridge: Polity.
- Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In *The Routledge Handbook of Corpus Linguistics* (pp. 14-28). Routledge.  
<https://doi.org/10.4324/9780203856949-2>
- Bradley, J. (2018). Digital tools in the humanities: Some fundamental provocations? *Digital Scholarship in the Humanities*, 34(1), 13-20.  
<https://doi.org/10.1093/llc/fqy033>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage Publications.
- Brown, A., & Tor, B. (2023, January 27). *Library guides: Tutorials + resources: Voyant*. Library Guides at the University of California, Santa Cruz. Retrieved from:  
<https://guides.library.ucsc.edu/DS/Resources/Voyant#:~:text=Researchers%20have%20used%20Voyant%20Tools,documentation%20do%20not%20>
- Brown, R. B. (2006). *Doing your dissertation in business and management: The reality of research and writing*. Sage Publications.
- Burdick, A., Drucker, J., Lunenefeld, P., Presner, T., & Schnapp, J. (2012). *Digital humanities* (1st ed.). MA: MIT Press.
- Burrell, G., & Morgan, G. (2016). *Sociological paradigms and organisational analysis*. Abingdon: Routledge.
- Burrows, J. (2002). Delta: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267-287.  
<https://doi.org/10.1093/llc/17.3.267>
- Byjus. (2019, May 21). *Taylor principles of scientific management: Meaning, definition*. Byjus. Retrieved October 8, 2023, from <https://byjus.com/commerce/taylor-principles-of-scientific-management/>
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining. From Concept to Implementation*. NJ: Upper Saddle River.

- Cabena, P., Hadjnia, P., Stadler, R., Verhees, J., Zanasi, A., & International Business Machines Corporation. (1998). *Discovering data mining: From concept to implementation*. Upper Saddle River, NJ: Prentice Hall.
- Castellà, Q., & Sutton, C. (2014). Word storms: Multiples of word clouds for visual comparison of documents. *International World Wide Web Conference Committee (IW3C2)*. Retrieved November 2, 2021, from [homepages.inf.ed.ac.uk/scutton/publications/castella14word.pdf](http://homepages.inf.ed.ac.uk/scutton/publications/castella14word.pdf)
- Caulfield, J. (2019, September 6). *How to Do Thematic Analysis | A Step-by-Step Guide & Examples*. Scribbr. Retrieved November 2, 2021, from <https://www.scribbr.com/methodology/thematic-analysis/>
- Chang, F. H. (2014). *Teacher education policies and programs in Pakistan: The growth of market approaches and their impact on the implementation and the effectiveness of traditional teacher education programs* [Doctoral dissertation]. [https://d.lib.msu.edu/islandora/object/etd:2681/datastream/OBJ/download/Teacher\\_education\\_policies\\_and\\_programs\\_in\\_Pakistan\\_the\\_growth\\_of\\_market\\_approaches\\_and\\_their\\_impact\\_on\\_the\\_implementation\\_and\\_the\\_effectiveness\\_of\\_traditional\\_teacher\\_education\\_programs.pdf](https://d.lib.msu.edu/islandora/object/etd:2681/datastream/OBJ/download/Teacher_education_policies_and_programs_in_Pakistan_the_growth_of_market_approaches_and_their_impact_on_the_implementation_and_the_effectiveness_of_traditional_teacher_education_programs.pdf)
- Clouder, L., & King, V. (2015). What works? A critique of appreciative inquiry as a research methodology. *Theory and Method in Higher Education Research*, 169-190. <https://doi.org/10.1108/s2056-375220150000001008>
- Cohen, L., Manion, L., Morrison, K., & Morrison, R. B. (2007). *Research Methods in Education*. Routledge.
- Craig, D. H., & Kinney, A. F. (2009). *Shakespeare, computers, and the mystery of authorship*. Cambridge University Press.
- Craig, H. (2010). Style, statistics, and new models of authorship. *Early Modern Literary Studies*, 15(1).
- Craig, H. (2011). Shakespeare's vocabulary: Myth and reality. *Shakespeare Quarterly*, 62(1), 53-74. <https://doi.org/10.1353/shq.2011.0002>
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage
- Cristobal, R., & Sebastian, V. (2013). *Data mining in education*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.

- Culpeper, J. (2002). Computers, language and characterisation: An analysis of six characters in *Romeo and Julie*. Retrieved from [https://lexically.net/wordsmith/corpus\\_linguistics\\_links/Keywords-Culpeper.pdf](https://lexically.net/wordsmith/corpus_linguistics_links/Keywords-Culpeper.pdf)
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Julie*. *International Journal of Corpus Linguistics*, 14(1), 29-59.
- Daines, J. G. III, Nimer, C. L., & Lee, J. R. (2018). Exploring the American Archivist: Corpus analysis tools and the professional literature. *Journal of Contemporary Archival Studies*, 5(3), 1-18.
- Danino, M. (2010). *The lost river: On the trail of the Sarasvati*. Penguin Books India.
- Darwesh, M. A. (2020). *Motif And Mode: An Interactive Word Cloud Approach to Tagore's Gitanjali* [Unpublished master's thesis]. National University of Modern Languages (NUML), Islamabad, Pakistan.
- Dickerson, M., & Pack, D. (2014). A Gentle Introduction to Text Analysis with Voyant Tools. *Research Librarian for Digital Humanities and History*, UCI Libraries.
- Driessen, K. (2021, March 16). *What is the difference between a longitudinal study and a cross-sectional study?* Scribbr. Retrieved February 29, 2023, from <https://www.scribbr.com/frequently-asked-questions/longitudinal-study-vs-cross-sectional-study/#:~:text=Longitudinal%20studies%20and%20cross%2Dsectional,an%20extended%20period%20of%20time>
- Drouin, J. (2014). Close- and Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching the *Little Review*. *The Journal of Modern Periodical Studies*, 5(1), 110-135. <https://doi.org/10.5325/jmodeperistud.5.1.0110>
- Durant, G. B. (2004). A typology of research methods within the social sciences. *NCRM Working Paper*, 1-22. Retrieved from <http://eprints.ncrm.ac.uk/115/>
- Eder, M., Piasecki, M., & Walkowiak, T. (2017). An open stylometric system based on multilevel text analysis. *Cognitive Studies | Études Cognitives*, 2017(17). <https://doi.org/10.11649/cs.1430>
- Edison, T. A. (2001). *Brainy Quote*. Retrieved May 6, 2018, from [https://www.brainyquote.com/quotes/thomas\\_a\\_edison\\_136633](https://www.brainyquote.com/quotes/thomas_a_edison_136633)
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. AAAI (American Association for Artificial Intelligence) Press.

- Fink, A. (2019). *Conducting research literature reviews: From the internet to paper* (6th ed.). SAGE Publications.
- Flick, U. (2014). *The SAGE handbook of qualitative data analysis*. NJ: SAGE
- Fowler, A. (2014, October 1). *What's in a literary name?* OUPblog. Retrieved March 29, 2023, from <https://blog.oup.com/2012/09/whats-in-a-literary-name/>
- Gao, W., & Wallace, L. (2017). Data mining, visualizing, and analyzing faculty thematic relationships for research support and collection analysis. *ACRL 2017 Conference Proceedings*. Baltimore: ALA, 171-178.
- Gee, J. P. (2012). *Situated Language and Learning: A Critique of Traditional Schooling*. Florence: Taylor and Francis.
- Gibbs, G. R. (2012). Thematic coding and categorizing. In *Analyzing Qualitative Data* (pp. 1-22). London: SAGE Publications, Ltd.  
[https://www.betterevaluation.org/sites/default/files/Thematic\\_coding\\_and\\_categorizing.pdf](https://www.betterevaluation.org/sites/default/files/Thematic_coding_and_categorizing.pdf)
- Godwin-Jones, R. (2006). Tag clouds in the blogosphere: Electronic literacy and social networking. *Language, Learning and Technology*, 10(2), 8-15.  
<http://dx.doi.org/10125/44055>
- Gregory, K., Geiger, L., & Salisbury, P. (2022). Voyant tools and descriptive metadata: A case study in how automation can compliment expertise knowledge. *Journal of Library Metadata*, 22(1-2), 1-16. <https://doi.org/10.1080/19386389.2022.2030635>
- Grier, M. (2014, October 29). *Voyant Tools | DH and Literary Studies*. Retrieved on February 12, 2022, from <http://modernist-magazines.org/?q=taxonomy/term/782>
- Haigh, T. (2019). *Exploring the early digital*. Springer.
- Hayes, S. (2008). Toolkit: Wordle. *Voices from the Middle*, 16(2), 66-68.
- Hearst, M. (2003, October 17). *What is text mining?* Berkeley. Retrieved March 31, 2023, from <https://people.ischool.berkeley.edu/~hearst/text-mining.html>
- Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3-10). Association for Computational Linguistics.
- Hendrigan, H. (2019). Mixing digital humanities and applied science librarianship: Using Voyant tools to reveal word patterns in faculty research. *Issues in Science and Technology Librarianship*, 2019(91), 21-32. <https://doi.org/10.29173/istl3>
- 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>
- Heydenreich, L. H. (2019, January 25). *Leonardo da Vinci | biography, art, & facts*. Retrieved February 2, 2022, from <https://www.britannica.com/biography/Leonardo-da-Vinci>
- Hirsch, B. D. (2012). *Digital humanities pedagogy: Practices, principles and politics*. Open Book Publishers.
- Hoover, D. L., Culpeper, J., & O'Halloran, K. (2014). *Digital literary studies: Corpus approaches to poetry, prose, and drama*. Routledge.
- Hope, J., & Witmore, M. (2004). The very large textual object: A prosthetic reading of Shakespeare. *Early Modern Literary Studies*, 9(3), 1-36.
- Hope, J., & Witmore, M. (2010). The hundredth psalm to the tune of *Green Sleeves*: Digital approaches to Shakespeare's language of genre. *Shakespeare Quarterly*, 61(3), 357-390. <https://doi.org/10.1353/shq.2010.0002>
- IBM. (2023, February). *What is text mining?* IBM - US. Retrieved May 27, 2023, from <https://www.ibm.com/topics/text-mining#:~:text=the%20next%20step-,What%20is%20text%20mining%3F,meaningful%20patterns%20and%20new%20insights>
- Ignatow, G. (2015). Theoretical Foundations for Digital Text Analysis. *Journal for the Theory of Social Behaviour*, 46(1), 104-120. <https://doi.org/10.1111/jtsb.12086>
- Jameslopresti. (2022, June 28). *Scientific management theory: Definition, history, principles, goals*. Villanova University. Retrieved October 8, 2023, from <https://www.villanovau.com/articles/leadership/scientific-management-theory-explained/>
- Jensen, K. E. (2014). Linguistics in the digital humanities: (computational) corpus linguistics. *MedieKultur: Journal of media and communication research*, 30(57), 115-134. <https://doi.org/10.7146/mediekultur.v30i57.15968>
- Jesson, J. (2011). *Doing your literature review: Traditional and systematic techniques*. SAGE.
- Jockers, M. L., & Underwood, T. (2015). Text-mining the humanities. In *A New Companion to Digital Humanities* (pp. 291-306). <https://doi.org/10.1002/9781118680605.ch20>
- John. (2017, August 20). *How to write a methodology*. Dissertation Help. Retrieved September 13, 2021, from <https://writepass.com/journal/2017/08/write-methodology-dissertation-help/>



- Jones, I. (2022). Research questions, aims and objectives. In *Research Methods for Sports Studies* (4th ed., pp. 47-61). Routledge. <https://doi.org/10.4324/9781003195467-4>
- Jones, S. E. (2013). *The emergence of the digital humanities*. Routledge.
- Kairaitytė-Užupė, A., Ramanaukaitė, E., & Rudžionis, V. E. (2023). Scientific information analysis using text analysis tool “Voyant tools”. *Information & Media*, 97, 25-48. <https://doi.org/10.15388/im.2023.97.57>
- Kelemen, M., & Rumens, N. (2008). *An introduction to critical management research*. London: Sage.
- Kemman, M. (2016, November 24). *A Republic of Emails: What are the contents?*. Retrieved September 22, 2021, from <https://www.maxkemman.nl/2016/11/a-republic-of-emails-what-are-the-contents/>
- Khan, M. Y., & Tufail, H. (2020). An Investigation into the Effectiveness of MALL during COVID-19 at the Higher Education in Pakistani EFL Classrooms. *Global Language Review*, 1(1), 175-185. [https://doi.org/10.31703/glr.2020\(V-I\).19](https://doi.org/10.31703/glr.2020(V-I).19)
- Kilgarriff, A. (2014, March 12). *Corpora in English language teaching*. British Council. Retrieved February 17, 2023, from <https://www.britishcouncil.org/voices-magazine/corpora-english-language-teaching>
- Kiselev, A. (2021). Using voyant tools with Japanese YouTube comments. *The European Journal of Humanities and Social Sciences*, (4), 38-43. <https://doi.org/10.29013/ejhss-21-4-38-43>
- Koehler, M. J., & Mishra, p. (2009). What Is Technological Pedagogical Content Knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60-70.
- Kolb, D. A. (1981). Learning styles and disciplinary differences. *The modern American college*, 232-255.
- Krippendorff, K. (2003). *Content analysis: An introduction to its methodology*. SAGE.
- Kumar, R. (2020). *Research methodology: A step-by-step guide for beginners*. SAGE.
- Lankshear, C., & Knobel, M. (2011). *New literacies: Everyday practices and classroom learning*. Maidenhead: Open University Press
- Lazer, D. (2008). Life in the network— the coming age of computational social science. In *CRCS Privacy and Security Lunch Seminar*.
- Lessa, B. (2020). Análisis con Voyant tools del primer numero de la Revista de Humanidades Digitales. *Publicaciones de la Asociación Argentina de Humanidades Digitales*, 1, e012. <https://doi.org/10.24215/27187470e012>

- Levenberg, L., Neilson, T., & Rheams, D. (2018). *Research methods for the digital humanities*. Springer.
- Li, Y., Ji, W., & Xu, D. (2017). Quantitative style analysis of Mo Yan and Zhang Wei's novels. *Proceedings of the International Conference on Web Intelligence*.  
<https://doi.org/10.1145/3106426.3109045>
- Maci, S. M., & Sala, M. (2022). *Corpus linguistics and translation tools for digital humanities: Research methods and applications*. Bloomsbury Publishing.
- Masood, M. H. (2022). *Study of Shakespearean themes and characters through corpus based Voyant tools approach* [Unpublished master's thesis]. University of Management of Technology (UMT).
- McCarthy, M., & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In *The Routledge handbook of corpus linguistics* (pp. 1-13). Routledge. <https://doi.org/10.4324/9780203856949-2>
- McCombes, S. (2023, January 2). *How to write a literature review | Guide, examples, & templates*. Scribbr. Retrieved April 29, 2023, from  
<https://www.scribbr.com/methodology/literature-review/>
- McGowan, B. S. (2021). Using text mining tools to inform search term generation: An introduction for librarians portal. *Libraries and the Academy*, 21(3), 603-618.  
<https://doi.org/10.1353/pla.2021.0032>
- McNaught, C., & Lam, P. (2014). Using Wordle as a supplementary research tool. *The Qualitative Report*, 15(3), 630-643. <https://doi.org/10.46743/2160-3715/2010.1167>
- Miguel, N. M. (2022). Using Voyant tools to enhance learners' reflections on their writing. In *Teaching English with Corpora* (pp. 155-157). Routledge.  
<https://doi.org/10.4324/b22833-32>
- Miley, F., & Read, A. (2011). Using word clouds to develop proactive learners. *Journal of the Scholarship of Teaching and Learning*, 11(2), 91-110.
- Miller, A. (2018). Text mining digital humanities projects: Assessing content analysis capabilities of Voyant tools. *Journal of Web Librarianship*, 12(3), 169-197.  
<https://doi.org/10.1080/19322909.2018.1479673>
- Mireille, R., & Culpeper, J. (2011). A new kind of dictionary for Shakespeare's plays: An immodest proposal. In *Stylistics and Shakespeare: Transdisciplinary Approaches* (1st ed., pp. 58-83). Cambridge University Press.

- Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4), 730-741.  
<https://doi.org/10.1016/j.dss.2012.05.030>
- Mountain, A., & Marshall, H. (2019, April 11). *Thematic Coding*. Better Evaluation. Retrieved November 2, 2021, from  
<https://www.betterevaluation.org/en/evaluation-options/thematiccoding>
- Neha, E. K. (2023). Identifying content-related and non-content-related queries in online discussion forums using Voyant tools. *International Journal of Information and Education Technology*, 13(4), 614-620.  
<https://doi.org/10.18178/ijiet.2023.13.4.1845>
- Nilsson, J., Sundberg, D., & Höglund, J. (2018). *Papa Revisited: A Corpus-Stylistic Perspective on the Style and Gender Representation of Ernest Hemingway's Fiction* [Master's Thesis]. Linnaeus University, Sweden, Europe. <http://nu.diva-portal.org/smash/record.jsf?pid=diva2%3A1175700&dswid=7129>
- Nyhan, J., Rockwell, G., Sinclair, S., & Ortolja-Baird, A. (2023). *On making in the digital humanities: The scholarship of digital humanities development in honour of John Bradley*. UCL Press.
- Petersen, L. B. (2010). *Shakespeare's errant texts: Textual form and linguistic style in Shakespearean 'Bad' quartos and Co-authored plays*. Cambridge University Press.
- Platte, B. (2017, March 30). *Text analysis using Voyant tools*. Retrieved May 24, 2021, from <https://blogs.reed.edu/ed-tech/2017/03/text-analysis-using-voyant-tools/>
- Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the horizon*, 9(5), 1-6.  
<https://doi.org/10.4135/9781483387765.n6>
- Prensky, M. (2012). *From digital natives to digital wisdom: Hopeful essays for 21<sup>st</sup> century learning*. Thousand Oaks, CA: Corwin
- Rambsy, K. (2016). Text-Mining Short Fiction by Zora Neale Hurston and Richard Wright using Voyant Tools. *Digital Humanities*, 251-258.
- Ridley, D. (2012). *The literature review: A step-by-step guide for students*. SAGE.
- Rockwell, G. (2003). What is Text Analysis, Really? *Literary and Linguistic Computing*, 18(2), 209-219. <https://doi.org/10.1093/lc/18.2.209>
- Rockwell, G. (2016). Is humanities computing an academic discipline? In *Defining digital humanities: A reader* (pp. 13-34). London: Routledge.

- Rockwell, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities*. MIT Press.
- Romero, C., Ventura, S., Delgado, J.A., De Bra, P. (2007). Personalized Links Recommendation Based on Data Mining. In *Adaptive Educational Hypermedia Systems*, to appear in: *Proceedings of the second European Conference on Technology Enhanced Learning, EC-TEL 2007*, Springer LNCS, Crete.
- Rudestam, K. E., & Newton, R. R. (2022, August 29). *Literature review*. University of Edinburgh. Retrieved April 29, 2023, from <https://www.ed.ac.uk/institute-academic-development/study-hub/learning-resources/literature-review#:~:text=A%20literature%20review%20is%20a,rather%20than%20a%20literature%20report>
- Sampsel, L. J. (2018). Voyant tools. *Music Reference Services Quarterly*, 21(3), 153-157. <https://doi.org/10.1080/10588167.2018.1496754>
- Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (6th ed.). Pearson Education Limited.
- Schreibman, S., Siemens, R., & Unsworth, J. (2008). *A companion to digital humanities*. Oxford: Blackwell.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins Publishing.
- Scrivner, O., & Davis, J. (2017). Interactive Text Mining Suite: Data visualization for literary studies. *Visualization method*, 29-38. Retrieved May 29, 2017, from <http://ceur?ws.org/Vol-1786/scrivner.pdf>
- Sinclair, J. M. (2007). *Collocation reviewed*. Manuscript. Tuscan Word Centre, Italy.
- Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching* (Vol. 12). Amsterdam: John Benjamins Publishing.
- Sinclair, J.M. (1996). The search for units of meaning. *Textus*, 9(1), 75–106.
- Sinclair, S., & Rockwell, G. (2015a). *Examples Gallery | Voyant Tools Documentation*. Retrieved May 29, 2017, from <http://docs.voyant-tools.org/about/examples-gallery/>
- Sinclair, S., & Rockwell, G. (2015b). *Principles of Voyant Tools | Voyant Tools Documentation*. Retrieved May 29, 2017, from <http://docs.voyanttools.org/context/principles/370>
- Sinclair, S., & Rockwell, G. (2015c). *Tools Index | Voyant Tools Documentation*. Retrieved March 12, 2017, from <http://docs.voyant-tools.org/tools/>

- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (1999). Automatic authorship attribution. *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*. <https://doi.org/10.3115/977035.977057>
- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2000). Automatic text categorization in terms of genre and author. *Computational Linguistics*, 26(4), 471-495. <https://doi.org/10.1162/089120100750105920>
- Steiner, C. M., Agosti, M., Sweetnam, M. S., Hillemann, E., Orio, N., Ponchia, C., ... Conlan, O. (2014). Evaluating a digital humanities research environment: The CULTURA approach. *International Journal on Digital Libraries*, 15(1), 53-70. <https://doi.org/10.1007/s00799-014-0127-x>
- Surbhi, S. (2016, December 8). *Difference Between Research Method and Research Methodology (with Comparison Chart)- Key Differences*. Retrieved August 9, 2017, from <http://keydifferences.com/difference-between-research-method-and-research-methodology.html>
- Szudarski, P. (2023). *Collocations, corpora and language learning*. Cambridge University Press.
- Taylor, F. W. (2016). *The principles of scientific management*. New York, NY: Cosimo Classics.
- Tokarz, R. E. (2017). Identifying criminal justice faculty research interests using Voyant and NVivo. *Behavioral & Social Sciences Librarian*, 36(3), 113-121. <https://doi.org/10.1080/01639269.2017.1771034>
- Tools, M. (2023, March 13). *MindTools | Home*. Develop your personal wellbeing and career skills. Retrieved October 7, 2023, from <https://www.mindtools.com/anx8725/frederick-taylor-and-scientific-management>
- Torre, S. G. (2019). Análisis de corpus con Voyant tools. *Programming Historian en español*, 3(3). <https://doi.org/10.46430/phes0043>
- Uboldi, G., & Caviglia, G. (2014). Information visualizations and interfaces in the humanities. *New Challenges for Data Design*, 207-218. [https://doi.org/10.1007/978-1-4471-6596-5\\_11](https://doi.org/10.1007/978-1-4471-6596-5_11)
- Ullah, Z. (2022). *Unveiling knowledge patterns in intermediate English textbooks through Voyant text mining tools: A digital humanities study* [Unpublished doctoral dissertation]. National University Of Modern Languages, Islamabad, Pakistan.
- 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. *Journal of Research in Social Sciences (JRSS)*, 7(2), 83-98.  
<https://jrss.numl.edu.pk/index.php/jrss/article/view/80>
- van Halteren, H., Baayen, H., Tweedie, F., Haverkort, M., & Neijt, A. (2005). New machine learning methods demonstrate the existence of a human Stylome. *Journal of Quantitative Linguistics*, 12(1), 65-77.  
<https://doi.org/10.1080/09296170500055350>
- Vanchena, L. A. (2012). Reading German Culture, 1789–1918. *Journal of Literary Theory*, 1-11. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0222-002276>
- Verhoeven, B., & Daelemans, W. (2014). CLiPS Stylometry Investigation (CSI) corpus: A Dutch corpus for the detection of age, gender, personality, sentiment and deception in text. *The 9th International Conference on Language Resources and Evaluation (LREC)*, 3081–3085.
- Walliman, N. S., & Walliman N. (2011). *Research methods: The basics*. Taylor & Francis.
- Wang, J., Zhao, J., Guo, S., North, C., & Ramakrishnan, N. (2020). ReCloud: Semantics-based word cloud visualization of user reviews. *Graphics Interface 2014*, 151-158.  
<https://doi.org/10.1201/9781003059325-20>
- Welsh, M. E. (2014). Review of Voyant Tools. *Collaborative Librarianship*, 6(2), 96-98.  
 Retrieved from  
<https://digitalcommons.du.edu/collaborativelibrarianship/vol6/iss2/8>
- White, D. (2008, July 23). *Not 'natives' & 'immigrants' but 'visitors' & 'residents'*.  
 Message posted to <http://tallblog.conted.ox.ac.uk>
- Wilks, Y. (2004). On the ownership of text. *Computers and the Humanities*, 38(2), 115-127. <https://doi.org/10.1023/b:chum.0000031184.28781.47>
- Williams, N. L., Inversini, A., Buhalis, D., & Ferdinand, N. (2015). Community crosstalk: An exploratory analysis of destination and festival eWOM on Twitter. *Journal of Marketing Management*, 31(9-10), 1113-1140.  
<https://doi.org/10.1080/0267257x.2015.1035308>
- Yeates, R. (2013, August 22). *Voyant tools*. Retrieved September 22, 2021, from  
<https://postapocalypticcities.wordpress.com/2013/05/02/voyanttools/>
- Zahedzadeh, G. (2017). Overt attacks and covert thoughts. *Aggression and Violent Behavior*, 36, 1-8. <https://doi.org/10.1016/j.avb.2017.06.009>
- Zhang, T., Damerau, F., & Johnson, D. (2001). Text chunking using regularized winnow. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL '01*. <https://doi.org/10.3115/1073012.1073081>