P-ISSN 2708-6453	Pakistan Languages and Humanities Review	April-June 2025, Vol. 9, No. 2
O-ISSN 2708-6461	https://doi.org/10.47205/plhr.2025(9-II)24	[285-306]

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

RESEARCH PAPER

Text Mining and Fuzzy Logic in Literary Analysis: A Computational Study of The Book of Dede Korkut Using Voyant Tools

[1]Huma Asif Malik* and [2]Dr.Summaira Sarfraz

1. Lecturer, Department of Humanities and Sciences, FAST-National University of Computer and Emerging Sciences, Lahore, Pakistan.
2. Professor, Department of Humanities and Sciences, FAST-National University of Computer and Emerging Sciences, Lahore, Pakistan.

***Corresponding Author** | ahuma1.malik@gmail.com

ABSTRACT

This study explores the computational and stylistic features of *The Book of Dede Korkut* using Voyant Tools and fuzzy logic. The research applies Corpus Linguistics and draws on Kamal Abdulla and Rafik Aliev's (2023) theoretical insights, highlighting how fuzzy logic, centered on uncertainty and symbolic reasoning, mirrors the non-linear nature of mythological storytelling. A corpus-based methodology was employed, analyzing sentence structures, co-occurrence patterns, and stylistic trends through visualization techniques such as word clouds and dispersion plots. Frequent terms included "logic" (239), "fuzzy" (203), "text" (112), "epos" (95), and "language" (94), with 3,213 unique word forms identified. The study demonstrates that combining text mining and fuzzy logic deepens the understanding of stylistic patterns and the complex organization of mythological narratives. Customized visual dashboards could improve VT's usability for complex corpora, while collaboration with linguists and computational scientists is key to refining the application of fuzzy logic in textual analysis and advancing computational studies of traditional narratives.

KEYWORDS Text Mining, Fuzzy Logic, Textual Patterns, Stylistic Framework, Voyant Tools, Visualization Techniques, Corpus-Based Methodologies

Introduction

This research paper presents a comprehensive stylistic analysis of the literary masterpiece 'The Book of Dede Korkut and Fuzzy Logic', a collection of epic folk tales from the Oghuz Turks, using computational methods. The study aims to contribute to a better understanding of the stylistic features embedded in the text and to develop a framework for the analysis of literary style that can be applied to corpus-based studies. Fuzzy logic is intended to solve problems by taking into account all relevant data and selecting the optimal course of action given the available input. (Varshney & Torra, 2023) It has already been effectively used in artificial intelligence, machine learning, business decision-making and medicine. The use of Fuzzy logic in the text renders to misinterpretation of implicit meaning of the text. For the deeper meaning, the writer often uses literary devices such as irony, hedges, metaphors etc. in the form of fuzzy logic which makes the understanding a complex task. Therefore, this study is an attempt to provide a deep examination of linguistic and literary elements within a structured framework of "The book of Dede Korkut" in context with Fuzzy logic for the development of computational text analysis for the first time. For this purpose, the research aims to develop a Corpus-based framework for computational stylistic analysis of the text and offering a visualization interpretation of the said book. The epic history of the Oghuz Turks is told in the legends of Dede Korkut.

After several centuries, they abandoned their shamanic customs and adopted Islam as they traveled across the steppes of what Rory Stewart referred to as "The Land In Between" and settled in Anatolia.

The morals and ideals conveyed in the stories are important to the pre-Islamic beliefs and social customs of the nomadic Turkic peoples. The mythological storyline of the book is a component of the cultural legacy of Oghuz people, who are primarily from Turkmenistan, Azerbaijan, and Turkey.

The writers Kamal Abdulla, Rafik Aliev of "The Book of Dede Korkut and Fuzzy Logic" discovers many new horizons by the scientific direction to review a literary text in the context of fuzzy logic. It indicates a possibility that the principles of fuzzy logic already existed in the mental and language corpus of ancient Oghuz stories. It suggests the antiquity of the text to its affiliation with mythological history. The approach also provides an understanding the clear interrelation between language and consciousness.

Multiple possible truth values can be processed through a single variable when using fuzzy logic as a variable processing method. Using algorithms and an open, imprecise spectrum of facts, fuzzy logic seeks to solve problems and produce a range of exact results.

Fuzzy logic is intended to solve problems by taking into account all relevant data and selecting the optimal course of action given the available input. It is a logic of doubts, hesitations, and maybes. It is a logic of searches. It is the logic of efforts and endeavors. It is the logic of exploration and discovery of new horizons. It is the logic of a race towards quantums.

It was in 1965 a work for the journal Information and Control that Lotfi Zadeh originally suggested fuzzy logic. In his work, "Fuzzy Sets," Zadeh aimed to derive the fundamental logical rules for this type of set-in order to represent the type of data that is utilized in information processing. Since then, fuzzy logic has been effectively used in artificial intelligence, machine learning, business decision-making, image processing, aircraft engineering, machine control systems, and automotive traffic control. The individual shares his personality in this direction by equipping with the characteristics associated with national elements, thus gaining a national identity (Ay and Güllü, 2020).

Literature Review

This section of the paper discusses the ideology and assumption of educational change and leadership**.** According to Horenberg, n.d. Voyant Tools (VT) has been recognized as a valuable resource in literary studies, particularly in non-automated Computer-Assisted Literary Translation (Horenberg, n.d.). Unlike fully automated translation technologies, VT provides an interactive and minimally invasive approach to source-text analysis, enabling researchers and translators to explore linguistic and stylistic patterns. Horenberg (n.d.) investigates its use with a case study of Evelyn Waugh's (1936) short tale Mr Loveday's Little Outing, illustrating how Voyant Tools enables a systematic and data-driven analysis of literary aspects prior to translation. Voyant Tools can be utilized to generate a systematic overview of the corpus to further investigate about word frequencies, recurring phrases and stylistic trends. In their study, Horenberg (n.d.) emphasizes that VT's employs quantitative linguistic analysis that helps scholars and translators in identification of important textual cues to employ computational techniques and giving insights otherwise overlooked during traditional close reading.

Miller, (2018) emphasized Voyant Tools (VT) is an efficient platform for well-organized and structured text analytics. He asserts that information extraction is important for analyzing and interpreting textual data as it allows scholars and researchers to efficiently discover interesting and valid patterns within large corpora and VT makes it possible. During the test run of the Trials and Triumphs project at Middle Tennessee State University, Voyant Tools was used to announce its possibility for text analysis and its future incorporation with Drupal for digital engagement. The research proved the software's efficacy in analyzing themes, connections and developments in historical texts by processing and visualizing language patterns. He also stated that Tool is a valuable addition for scholars researching digital humanities as it supports both quantitative and qualitative approaches for text mining. It amplifies scholars' understanding for corpora through interactive visualization and content analysis tools for word distributions, thematic formations, and stylistic traits.

Gregory et al. (2022), evaluates its potential for generating archive metadata, demonstrating how librarians, archivists, and catalogers can imply VT's systematic thinking to create significant and confirmable subject keywords from large textual data. Through OCR-ed and digitized text processing, VT helps in determining the thematic intent of collections and quality of metadata. The study lays importance on how data visualization tools in VT help users to discover trends, validate the relevance of content, and narrows the process of developing metadata. They further range over software's ability for professional judgment but does not substitute for human ability in order to facilitate more informed and systematic metadata creation. The way in which computational resources assist corpus analysis, linguistic documentation, and textual interpretation aligns with more meaningful and broader research areas in digital humanities.

Ullah, (2022) studies Voyant Tools by using five basic areas Summary, Cirrus, Phrases, Links, and Context to analyze mid-level English textbooks based in Pakistan. Transforming the textual content into dynamic visualizations, the tool enhances textbook engagement and provides effective and engaging reading experience. The Summary tool, findings provide insights that software can quantify stylometric features like frequency of words, vocabulary density, and sentence length, while Phrases and Cirrus indicate significant themes, collocations, and repetitive linguistic patterns (Ullah, 2022). The Links tool makes knowledge graphs by linking important concepts, while the Context tool implores word meanings by analyzing how words are distributed across different textual contexts. This study underscores the prospective of distant reading to enhance comprehension, independent learning, and corpus creation in educational and pedagogical settings.

Feldman & Sanger, 2007, emphasis VT's text mining application in identifying patterns and significant themes in natural language text. Joudrey & Taylor, 2017 explores resource management and access to information are enhanced by technology-based solutions in smart libraries through VT. Many studies indicate that text mining techniques such as n-grams, concordances and corpus analysis can be applied for discovering trends in themes in library sciences (Kumar & Singh, 2021). Voyant Tools has demonstrated potential in data visualization and content analysis as an open-source text-mining resource. (Sinclair & Rockwell, 2016). Through the employment of text mining tools to analyze smart library material, this research builds on prior research by addressing more detailed insight into themes and trends in large datasets.

Material and Methods

Computerized text analysis has seen tremendous growth over the past decades. This is mainly because of improvements in information technology and computer program development but also because high interest in applying electronic resources for assisting and utilizing more conventional methods of literary and linguistic research. Particularly within the context of higher education, increased computer availability has been a factor in the expanding acceptance of computer-based textual analysis. The development of organized bodies of electronic texts, or corpora, has enabled the structured and formed investigation of frequent patterns in language usage. Corpus linguistics, as a new area of research, has a pivotal role in this development. Adolph, S. (2006) The field of stylistic analysis has helped researchers to traditionally study the linguistic intricacies of writings to find their internal structure and hidden meanings.

Theoretical Framework

The stylistic text analysis measures the visualization and interpretation of a theoretic model and integrates theories of computational linguistics and cognitive processing. Systemic Functional Linguistics (SFL) by (Halliday, 1978) offers an extensive theoretical model for understanding choices in language within a social and communicative context. It focuses on how linguistic choices construct meaning, which can be assessed later using corpus-based approaches.

The stylistic patterns of the text involve Word clouds, graphs, and dispersion plots as some of the visualization tools. Visual representation of word distribution enables scholars to interpret word distribution, identify stylistic marks, and understand the general organization of the text (Liu, 2011). Corpus linguistics and computational stylistics provide a solid base for analyzing complex relationships between language and style.

The analysis would be capable of detecting and evaluating different stylistic elements, ranging from discourse-level elements to lexical and grammatical structures, through statistical methods and computer resources. The combination of annotation, preprocessing, and quantitative analysis gives a deeper understanding of authors' stylistic choices with a new vision of how language constructs tone, voice, and genre.

Data Collection and Sampling

The research analyzes the random sampling method since the literature consists of many long texts, therefore to ensure there is a representative sample, parts and excerpts are selected at random. This process would help avoid bias in the selection process. The data was obtained from a book entitled "The Book of Dede Korkut and Fuzzy Logic," authored in 2023 in Baku, Azerbaijan by Kamal Abdulla and Rafik Aliev. By fuzzy logic, it comprises epos, narratives, and folklore of the heroic deeds of the ancient ancestors of the Oghuz people, primarily from Turkmenistan, Azerbaijan, and Turkey. Since there are numerous extensive works in the literature, sections and extracts are chosen at random to guarantee a representative sampling. This method would assist in preventing selection bias.

A list of all sentences in the text and use a random number generator to select a predetermined number of sentences for analysis would be created. This method would provide a broad overview of stylistic elements across the text keeping in mind the sentence length and complexity; random selection might include very short or long sentences that could skew results. Analysis of the entire text and randomly select a set number of words

(e.g., every 50th word) to examine their usage, frequency, and stylistic impact. This method would highlight particular linguistic patterns and capture context effectively. Identification of key themes and motifs in the text and randomly select sections that relate to each theme. Extracting the themes of love, conflict, and identity from the literary text would ensure a balanced representation of different thematic elements in the analysis. The text contains dialogue, randomly select a mix of dialogue and narrative passages to analyze how style varies between these forms. This technique would highlight differences in linguistic features between spoken and written forms within the same text.

Representativeness of Corpus

A comprehensive and systematic analysis of the literary text would involve organization of corpus representation. Representing corpora in a literary analysis involves creating a structured and organized collection of texts or excerpts that can be analyzed for various linguistic and stylistic features of "The Book of Dede Korkut" and Fuzzy Logic.

Textual Structure

Chapters/Sections:

In Lieu of introduction The Elements of Fuzzy Logic in "The book of Dede Korkut"

Chapter1: "The construction" Material of Fuzziness in "The Book of Dede Korkut"

Chapter 2: The text of "The Book of Dede Korkut" and the constant "Doubt" of Fuzzy Logic

Chapter 3: The Text of "The Book of Dede Korkut" and "The Search for the third" through Fuzzy Logic

Chapter 4: "My Beauty, you are more beautiful than the beauty of the Beauties!"

Chapter 5: "The Book of Dede Korkut" In the Epistemological light of Fuzzy Logic

Chapter 6: The Facet of relations of Fuzzy Logic and "The Book of Dede Korkut"

In Lieu of Conclusion

Meta Data documentation:

Title and Author: Kamal Abdullah and Rafiq Aliev

Publication Details: Baku Azerbaijan 2023

Genre: Epic Literature, folklore

Corpus Size and Composition

Word Count: Corpus had 4529 entries with total frequency of 23,962 words.

Computational Extension: Graphical User Interface (GUI)

The application uses Voyant Tools to create a graphical user interface. It includes an upload area where users can input text or upload text. There's a "Reveal" button that triggers the analysis process. Voyant Tool bar chart for the Parts of Speech Distribution, relative frequency distribution, correlations, word lists, knots, links, dream Scape, looms,

mandala, terms radio, scatter plots for high frequency words, collocates, stream graph for raw frequency words and textual arc for corpus. Entity Identification with Voyant Tools to identify entities like people, locations, cardinals, national/religious/political, geopolitical, work of art.

Results

Text Preparation

In the first phase of study for Text Preparation, the book was converted into PDF and text document via internet through an online pdf converter tool https://www.sodapdf.com/txt-to-pdf/ and further divided into six chapters and assigned page numbers. The title of the book was created using the Canva app. For the purpose of data cleaning the literary content of book was converted into .txt file by using https://www.freeconvert.com/word-to-txt/download Link and all the punctuation marks and capitalization were removed to make it ready for pre-processing.

Data Cleaning and Pre processing

Data cleaning and Preprocessing refers to the initial set of operations applied to raw text data to convert it into a format that can be more easily analyzed computationally. As a foundation for computational software to properly evaluate language features such as lexical decisions, syntactic structures, and stylistic elements, these processes ensure that the text is in a standardized, clean, and organized form.

The book was converted into a.txt file for data cleaning, and letters were capitalized as well as punctuation removed to make it clean for pre-processing. By using the https://www.freeconvert.com/word-to-txt/download link, the book was converted into a.txt file for corpus creation. All the punctuation and capitalization were removed to make it ready for pre-processing.

Visualization and Interpretation

Tandel et al., 2019 states that Text mining is the process of analyzing texts in natural language and subsequently searching for patterns in the lexicon to retrieve useful information. With the inculcation of natural language processing (NLP) methods, it converts unstructured data into structured form of data for analysis and tries to retrieve useful information that is helpful for a particular purpose (Kumar, 2013). Voyant Tools was employed in this study to analyze text. It is a web-based text analysis, reading, and visualization platform that was launched in 2003 (Sampsel, 2018). It was created by a small group of digital humanities researchers headed by Stefan Sinclair and Geoffrey Rockwell (S. G Sinclair, 2021). Voyant Tool is an open-access "web-based reading and analysis environment for digital texts" (Sinclair & Rockwell, 2020). It is an open-source suite of 20+ interactive instruments, enabling data extraction and offering textual as well as statistical analysis to reveal meaningful patterns or trends within or between texts (Jocelyn, 2020). Voyant Tool's tagline is, "see through your text" and permits to represent large corpus in a number of various manners (Sampsel, 2018). Voyant, allows multiple documents to be uploaded together and is called a “corpus.” Once it is uploaded, the research analysis can begin by selecting the different tools, also called “skins” (Sinclair, 2015). The default skin shows five different tools like Cirrus, Summary, Trends, Reader, and Contexts, but the analysis is not limited to just those tools, and the tools can be changed within the skin that configure data in various visualizations. Visualization and interpretation of corpus by Voyant Tools for CIRUS word clouds of high frequency words, trends, entity identification

count of person, product, date, cardinal, geo political entities, quantity, money, national, religious of digital text.

Summary Tool

This tool provides information about the total number of documents and words in a corpus. It also gives information about unique words, longest and shortest documents by the number of words in the corpus, top and the lowest vocabulary densities, average number of words per sentence (highest and lowest) and five most frequent words in the corpus. Table below provides summarized information about corpus.

This corpus has 9 documents with 23,962 total words and 3,213 unique word forms.

Document Length

Longest: page 22 (5924); Text-1 (4744); page 50 (2550); page 32 (2387); page 27 (2025)

Shortest: page 35 (1273); in lieu of introduction (1450); page 44 (1649); page 40 (1960); page 27 (2025)

Vocabulary Density:

Highest: page 35 (0.434); in lieu of introduction (0.388); page 32 (0.350); page 40 (0.345); page 44 (0.341)

Lowest: page 22 (0.254); Text-1 (0.265); page 50 (0.289); page 27 (0.340); page 44 (0.341)

Average Words Per Sentence

Highest: page 40 (23.3); page 27 (19.5); page 50 (19.0); Text-1 (18.9); page 22 (18.5)

Lowest: page 32 (16.6); page 35 (17.2); in lieu of introduction (17.5); page 44 (18.3); page 22 (18.5)

Readability Index:

Highest: page 40 (11.234); in lieu of introduction (11.114); page 35 (11.111); Text-1 (10.921); page 27 (10.464)

Lowest: page 50 (9.362); page 32 (9.753); page 22 (10.211); page 44 (10.274); page 27 (10.464)

Most frequent words in the corpus:

logic (239); fuzzy (203); text (112); epos (95); language (94)

Distinctive words (compared to the rest of the corpus):

Text-1: characters (15), psychological (9), facet (11), zadeh (15), multiset (8).

in lieu of introduction: green (7), zadeh (10), aristotle (5), meadow (3), ability (3).

page 22: characters (17), beyrek (22), beyrek's (10), gã (9), facet (22).

page 27: wife (14), dirse (16), wounded (5), khan's (10), kazilik (5).

page 32: kimi (15), hedge (21), pronoun (16), hedges (19), variable (8).

page 35: dönə (10), reduplicative (11), hedge (7), thousand (4), bin (4).

page 40: doubt (25), constant (11), absolute (9), voice (3), systems (3).

page 44: days (8), disease (4), period (11), doubt (7), death (7).

page 50: resurrect (4), member (4), wife (5), death (9), judgements (3).

Data Analysis through Voyant Tools

Visualization tools such as word clouds**,** graphs, and dispersion plots offer a visual representation of stylistic trends in the text. These tools help researchers interpret the distribution of words, identify stylistic markers, and understand the overall organization of the text in an intuitive, visual format (Liu, 2011). Incorporating the Corpus Linguistics Model visualization tools would help the in-depth analysis in corpus by highlighting the relationships between the top words, showing how frequently they co-occur or are connected. It provides insights into word associations and contextual relationships, helping users identify key themes or patterns in the text. KWIC, Word Clouds, ENTITY RECOGNITION, BUBBLELINES, LINKS, Trends, Root Word Generation of high frequency words, BAR Charts of relative and high frequency words, LOOMS, Scatter plots for high frequency words, Correlations, Dream Scape, Knots, MANDALA, Stream Graph for Raw Frequency words, Terms Radio Chart, Textual Arc for Corpus, are generated via VOYANT TOOLS. The KWIC tool generates a list of all instances of a search term in a corpus in the form of a concordance. The GraphColl tool identifies collocations and displays them in a table and as a collocation graph or network. The Text tool enables an in-depth insight into the context in which a word or phrase is used.

Word Tree Generation by Voyant Tools

A word tree in Voyant Tools shows the context of the most frequent keywords in a corpus by branching out with words or phrases that commonly appear before or after the keyword. The graphical representation aids in examining a text's recurrent language structures, relationships, and patterns (Heuser & Le-Khac, 2012). The word tree diagrams of some of the most common keywords in the corpus Fuzzy Logic, Logic, Language, and Korkut are displayed in the following images.

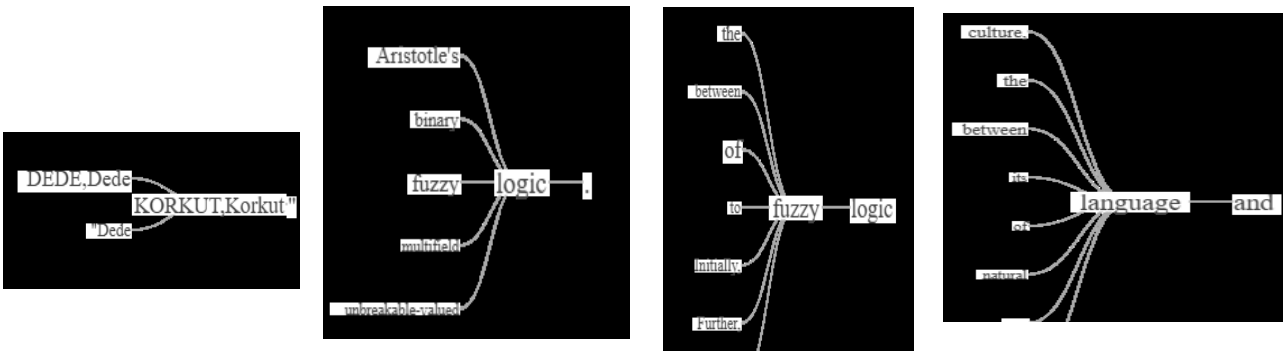

Fig.1 Root word generation of Korkut, Fuzzy Logic and Language

Trends in Data

Voyant Tools' Relative Frequency Stream Graph examines term usage over time or across texts by displaying the frequency of terms as a percentage of the overall word count within corpus sections. This form of visualization assists in the identification of theme shifts and word distribution trends in a text (Sinclair & Rockwell, 2016). The Relative Frequency Stream Graph is plotted below with respect to the corpus chapter-wise partition.

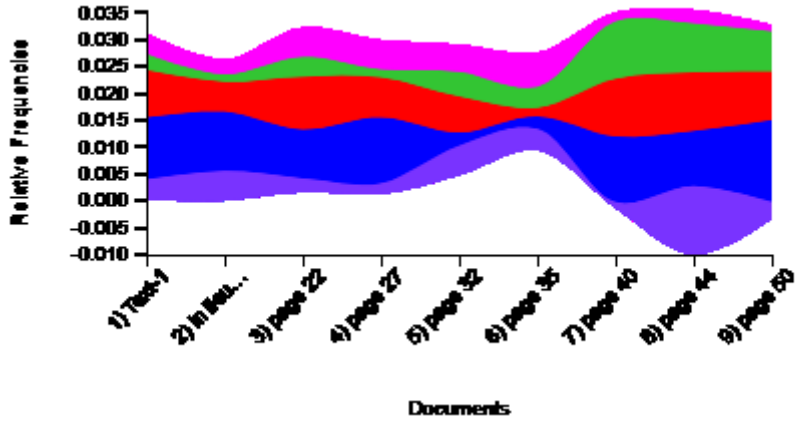

Fig.2 Relative Frequency Stream Graph

Voyant Tools Relative Frequency Line + Stacked Chart shows individual and cumulative patterns in the use of words by overlaying a line chart and a stacked area chart to show the relative frequency of a number of phrases throughout corpus regions. The graphic facilitates tracing trends in style and thematic evolution over time or between chapters (Sinclair & Rockwell, 2016). Using the corpus's chapter-wise distribution, the Relative Frequency Line Chart and Stacked Bar Chart are shown in the corresponding image.

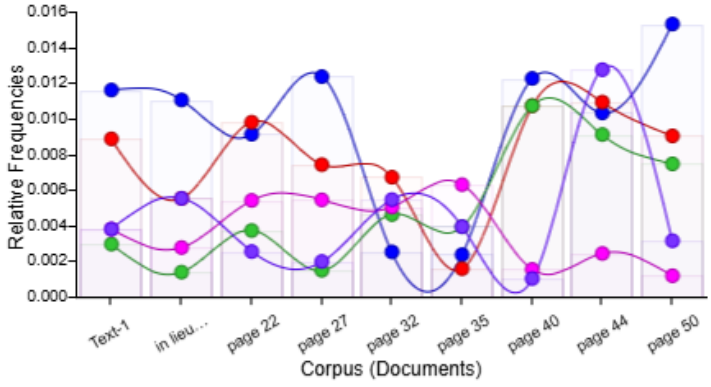

Fig.3 Relative Frequency Line+ Stacked Chart

Word usage within corpus can be contrasted due to Voyant Tools' Relative Frequency Bar Chart, which presents relative frequency of words as proportion bars. This tool helps systematic and comparative observation of important linguistic patterns, repeated themes, and stylistic differences by scholars (Sinclair & Rockwell, 2016). Referencing the corpus's chapter-by-chapter divide, the Relative Frequency Bar Chart is shown in the following graph.

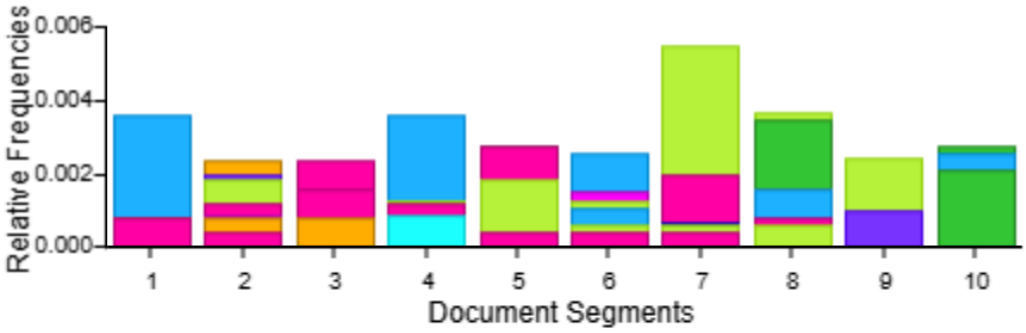

Fig.4 Relative Frequency Bar Chart

The Relative Frequency Line + Stacked Chart in Voyant Tools visualizes term trends over time, combining line and stacked area charts to show individual and cumulative frequencies. Following figure shows Relative Frequency Line Chart along with Stacked Chart with reference to the chapter wise division in corpus. Most frequent words were selected like fuzzy, logic, language, epos and language that gives information about the number of times these words show up in corpus. Each series in the graph is colored according to the word it represents. At the top of the graph a legend displays which words are associated with certain colors. Clicking on the words in the legend to toggle their visibility. Hovering over any point in the graph causes a callout box to appear with information about the point, including the word and the frequency.

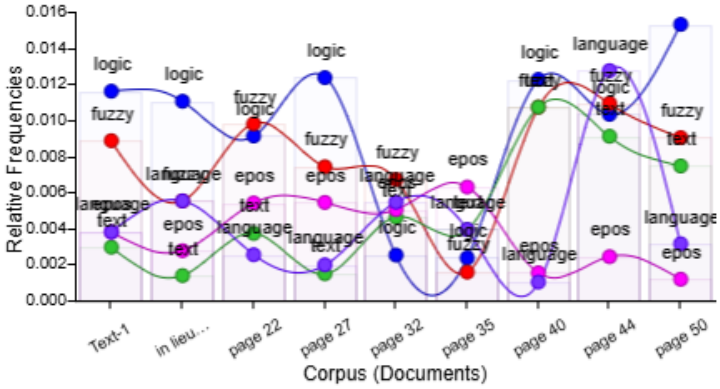

Fig.5 Relative Frequency Line + Stacked Chart

The Relative Frequency Column Chart in Voyant Tools displays the relative frequencies of terms as vertical bars, enabling comparison across documents or corpus sections. Following figure shows Relative Frequency Column Chart with reference to the chapter wise division in corpus.

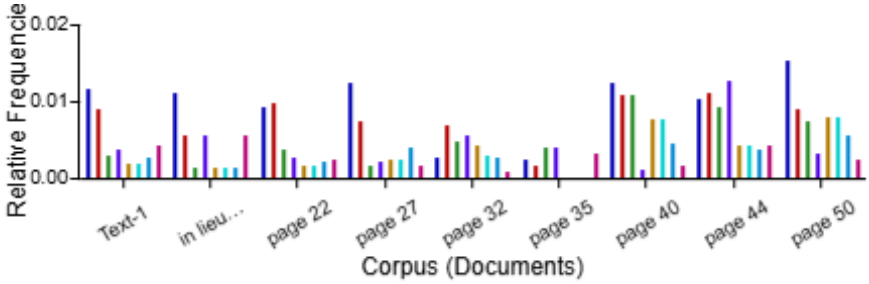

Fig.6 Relative Frequency Column Chart

This tool is highly beneficial for extracting linguistic elements from "The Book of Dede Korkut and Fuzzy Logic" by examining word distributions, thematic emphasis, and character references across several chapters. Researchers can investigate how fuzzy logic principles are mirrored in the text's storytelling strategies and language choices by monitoring the frequency of key terms associated with logic, mythological aspects, and narrative structures.

Cirus Word Cloud

Voyant Tools' Cirrus Word Cloud tool shows the most frequent terms in a corpus visually, with words that occur frequently showing up larger and more prominently. This tool facilitates the rapid identification of a text's main themes, lexical patterns, and prevalent vocabulary (Heuser & Le-Khac, 2012). The figure's correlations mostly show relative frequency, with larger terms denoting higher co-occurrence rates and better thematic relevance. Logic (239), fuzzy (203), text (112), epos (95), and language (94) are therefore commonly used words. Korkut (77), poles (77), Dede (74), book (71), world (67), and scheme (52) are next in line. The Cirrus Word Cloud is helpful in linguistic analysis for The Book of Dede Korkut and Fuzzy Logic because it emphasizes recurrent narrative themes, character names, and key ideas associated with the text's logical and mythological systems. Researchers might investigate how fuzzy logic concepts might be incorporated into the linguistic and thematic patterns of the text by graphically mapping the distribution of terms.

Fig.7 Cirus Word Cloud

Bubble lines

Bubble lines within Voyant Tools provide a graphically available means of examining data in a corpus through the dynamic visualization of phrase frequency patterns in documents. The "Trends" tool, which monitors term or keyword frequency in a corpus over time or in a number of publications, employs them as a visual representation (Sinclair & Rockwell, 2016). Each document in the corpus is represented as a horizontal line and divided into segments of equal length (50 segments by default). Each selected word is represented as a bubble with the size of the bubble indicating the word's frequency in the corresponding segment of text. The larger the bubble the more frequently the word occurs. Hovering over a location on the document line will cause a bubble to appear with term frequencies for that segment. At the end of the document line is a label that indicates the count of all terms selected for that document; hovering over the label shows a break-down of the term frequencies for that document. This visualization helps researchers in

analyzing linguistic trends and thematic movements by providing informative details regarding the corpus's document-based and time-based trends in word usage.

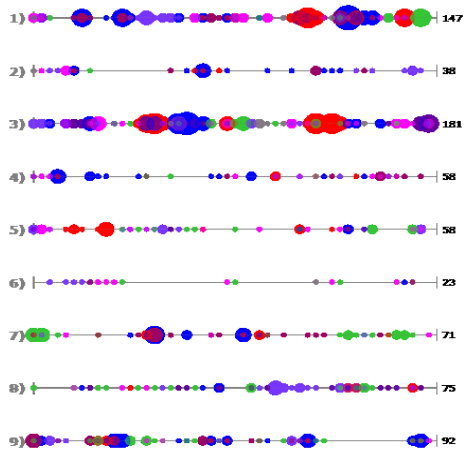

Fig. 8 Bubble Lines with reference to High Frequency Words in Corpus

Correlations

The Correlations tool enables an exploration of the extent to which term frequencies vary in sync (terms whose frequencies rise and fall together or inversely). They explain the relationship between a number of variables or characteristics of a corpus. Voyant has various table and visualization tools that can be utilized to analyze these relations and discover connections, patterns, and trends in the text (Sinclair & Rockwell, 2016). Co-occurrence and semantic co-occurrence are shown in the figure by putting terms that tend to appear together in the text in the same table sections. This comparison of the rate of occurrence of particular terms enables the discovery of underlying themes or correlations in the text. High frequency words could be a sign of the text's main ideas and themes. Researchers can gain deeper insights of the corpus's structure and significance by investigating correlations.

Correlations

Term 1	←	→	Term 2	Correlation...	Significanc...
language			thinking	-0.06458222	0.20490661
fuzzy			thinking	-0.06595065	0.21190907
epos			thinking	-0.06647721	0.19718572
great			logic	-0.0666816	0.6634091
text			thinking	-0.06793429	0.19289853
like			logic	-0.0795728...	0.644569
logic			world	-0.07980746	0.3127212
like			world	-0.084754	0.22467151
mindware			text	-0.09269884	0.3399817
epos			fuzzy	-0.09845615	0.4787611
language			world	-0.10296555	0.1488741
epos			world	-0.113718286	0.11922019
fuzzy			world	-0.116151825	0.13031343
text			world	-0.1213360...	0.10468502
like			text	-0.13097808	0.19627438
language			text	-0.17799023	0.09327088
epos			text	-0.2191052	0.049385604
logic			text	-0.23208855	0.06719954
fuzzy			text	-0.23884545	0.0433267

66 minimum coverage (%100) Scale

Fig.9 Correlations in Corpus

By offering interactive visualizations such as word clouds, correlation charts, and frequency charts that represent differences in diction, sentence structure, and thematic focus among the different parts of the text, functionalities such as Voyant Tools' Correlations feature enhance stylistic analysis. Through real-time tracking of shifts in tone, character dialogue, and plot development, these technologies facilitate comparative analysis. The method for observing how logic-based reasoning is embedded within the mythological structure of the text is to examine relative frequency patterns of fuzzy logic principle-related words compared to standard storytelling elements.

Dream Scape

Voyant Tools' Dream Scape is an interactive visualization tool that identifies theme patterns and contextual relationships between phrases by visualizing the interconnections and affinities among words in the corpus (Gregory, Geiger, & Salisbury, 2022). Through the assistance of this interactive and dynamic tool, one can visually analyze textual data and reveal obscured word correlations, highlighting crucial associations that may not be easily apparent. Dream Scape helps scholars recognize geographical and philosophical links within the book through references to places such as Plato in Brazil, Copenhagen, Kazan, Trabzon, and Bayburt by building contextual relationships between characters and their mobility. The suggestive patterns of recurring connections between locations, might help identify travel of people, ideas, goods, or anything else. The notion of travel here is to be interpreted loosely and critically: a sequence of locations, and Dreamscape seeks to help study them.

Through a rollover of words within the visualization provides a structured method of exploring language patterns and their connection to thematic frameworks in The Book of Dede Korkut and fuzzy logic. The recurring themes, geographical references, interpersonal interactions of characters, and logical conclusions can be incorporated within the text by charting relations among words and ideas. Through revealing patterns of repetition, contrast, and theme development within different parts of the corpus, it allows for a deeper and interesting understanding of how mythological stories work together with fuzzy logic concepts.

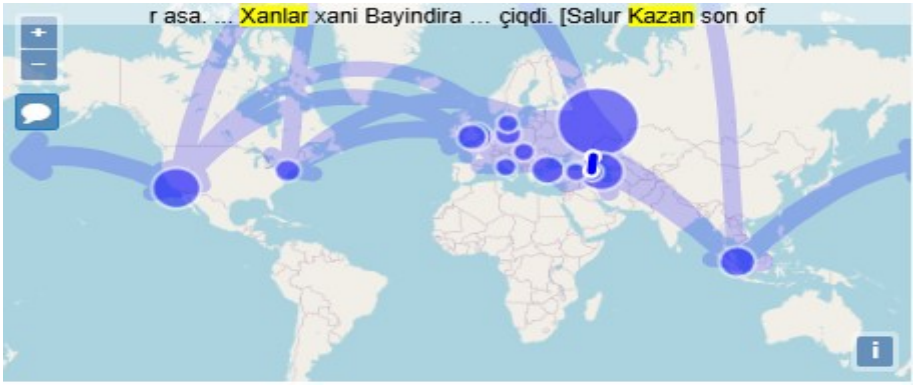

Fig.10 Dream Scape in Corpus

High-frequency terms in Corpus

A corpus's high-frequency words represents linguistic patterns, recurring themes, and frequent subjects. High-frequency terms that appear often across the corpus are shown in the image, while low-frequency terms are those that only appear once. Such words may be employed to discover the text's key ideas, character allusions, and stylistic tendencies (Miller, 2018). Researchers gain a deeper understanding about the lexical composition and

thematic emphasis of The Book of Dede Korkut and Fuzzy Logic through an analysis of the high- and low-frequency word distribution through this visual representation.

Fig.11 High frequency terms in Corpus

KNOTS in corpus

Voyant's color-coded Links visualization, connecting multiple knots to represent correlations among words, is useful for identifying themes, analyzing word connections, and analyzing the corpus's overall make-up (Gregory, Geiger, & Salisbury, 2022). Knots is a creative visualization that represents terms in a single document as a series of twisted lines. Each occurrence of a term is represented by a bend in the line, so the more twisted a line, the more a term repeats and straight stretches represent no occurrences. They are an important building block for showing the connections among words in a corpus and are utilized to depict words and ideas in the software. The relations among the knots are indicated by the lines connecting them, which usually depends upon contextual equivalence and co-occurrence in the corpus. The size and position of the knots are indicative of the frequency of specific terms, whereas the connections shown in Fig. 12 help reveal patterns of association and semantic relationships of logic (275); fuzzy (268); text (144); Korkut (119); Dede (116) with color coded link visualization.

Fig. 12 Representing patterns in Corpus through KNOTS

High frequency terms Links

Links are referred to as the visualization of most frequent words in a corpus as well as their connection to other terms. Co-occurrence relationships of high frequency words are represented by these links. A useful tool for corpus-based linguistic research, this kind of visualization aids in locating important topic structures, semantic linkages, and stylistic trends within the text (Miller, 2018).

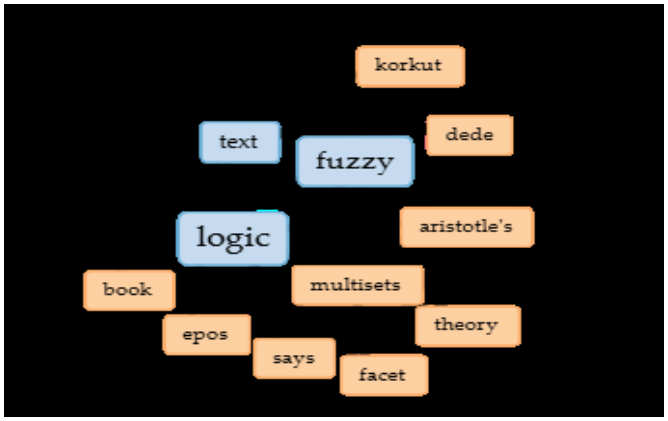

Fig.13 High frequency terms Links in Corpus

The links in Fig. 13are displayed as lines connecting terms (represented by knots), with thicker lines indicating stronger associations between the terms of Fuzzy, logic and text. This helps to identify key thematic connections and the semantic network of high-frequency words in the corpus. The figure above represents a network graph of the whole corpus where keywords in blue are shown linked to collocate in orange. Hovering over a term shows its frequency (for keywords it's the corpus frequency, for collocates it's the frequency in the context of the linked keywords). Here all words in blue are keywords like logic (n=239), fuzzy(n=203), text (n=112), and words in orange like multisets (n=21), Aristotle's (n=18) etc. are collocates.

High Frequency Terms Looms

High-Frequency Terms Looms in Voyant Tools refer to the most commonly occurring words or phrases in a corpus. These terms are often visualized in a word cloud or listed with their frequency count, providing insights into prominent themes and key topics within the text. Such visualizations help researchers analyze linguistic patterns and thematic structures efficiently (Gregory, Geiger, & Salisbury, 2022). Fig 14 shows a representation of high frequency terms Looms in corpus using Voyant Tools.

Fig.14 High frequency terms Looms in Corpus

Mandala in Corpus

In Voyant Tools, the Mandala visualization represents the distribution of terms in a circular layout. It shows words in a radial pattern, with the size of each segment reflecting

the frequency of the term in the corpus (Miller, 2008). Words are arranged around the circle, and their proximity to each other indicates semantic relationships or how often they co-occur in the text. In Fig.15, The Mandala visualization helps to provide a holistic view of word frequency and the interconnectivity of terms with mentioned page numbers of corpus in a visually striking way.

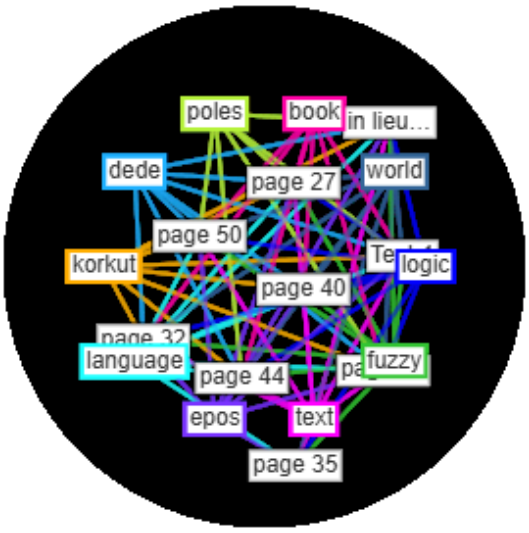

Fig.15 Mandala in Corpus

Scatter plot for High frequency words

The scatter plot in Voyant Tools visualizes high-frequency words by plotting their occurrences across documents, highlighting distribution patterns and trends. Following figure shows Relative Frequency Column Chart with reference to the chapter wise division in corpus.

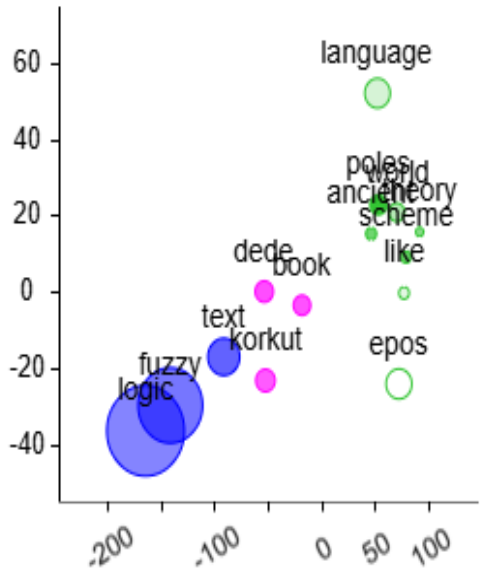

Fig. 16 Scatter plot for High frequency words in Corpus

Following figure shows Scatter plot for Raw frequency words with reference to the chapter wise division in corpus.

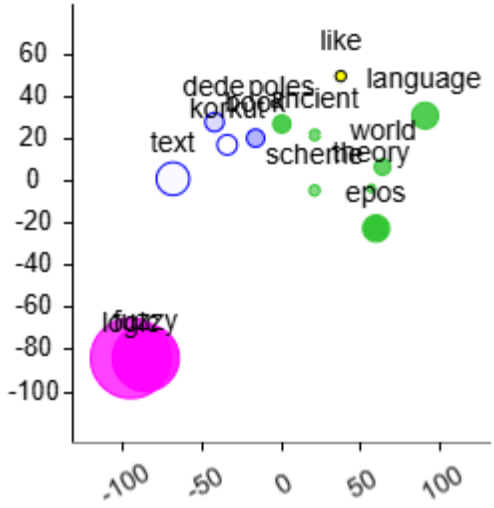

Fig.17 Scatter plot for Raw frequency words in Corpus

Terms Radio

The Terms Radio in Voyant Tools dynamically compares the frequency of selected terms across different documents or sections of the corpus, using a circular visualization. Terms Radio in Voyant Tools is interactive, allowing users to select terms and adjust settings to explore their distribution across the corpus. Following figure shows Terms Radio of frequently occurring with reference to the chapter wise division in corpus.

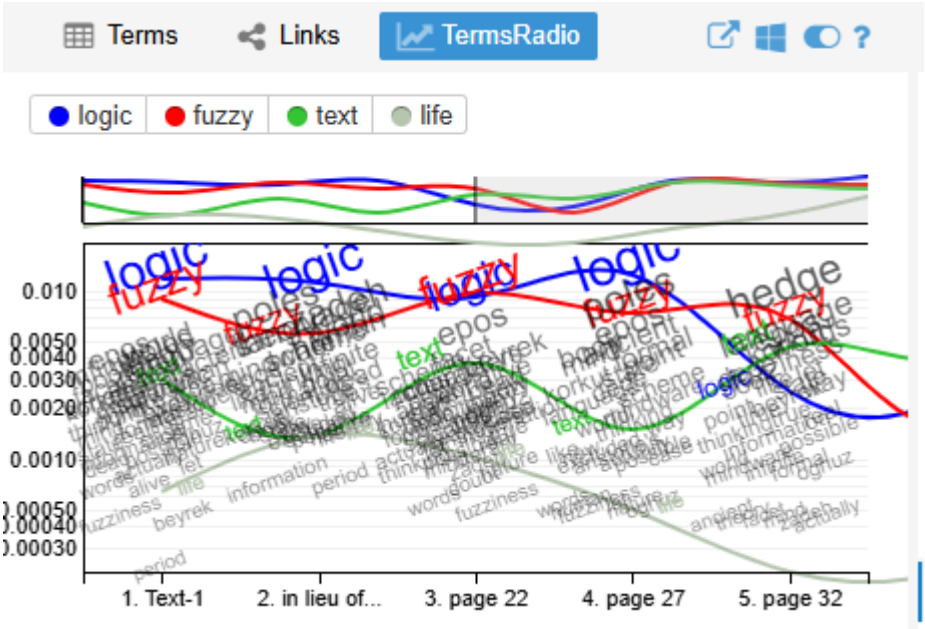

Fig. 18Terms Radio in Corpus

Stream Graph for High frequency words

The Stream Graph in Voyant Tools visualizes the relative frequency of high-frequency words over a corpus, with flowing, stacked layers representing term trends across sections or documents.

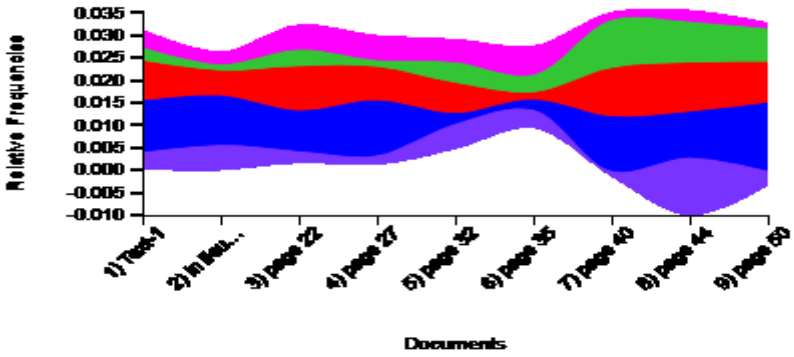

Fig. 19 Stream Graph for High frequency words in Corpus

The above figure shows Stream Graph for High frequency words with reference to the chapter wise division in corpus.

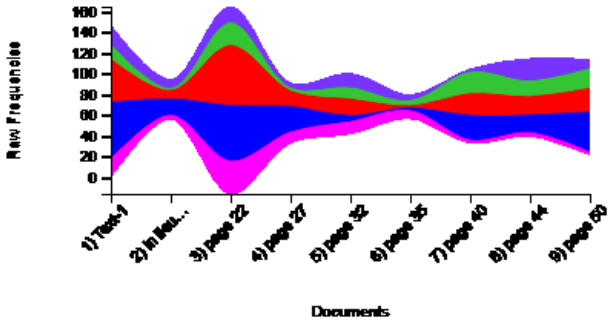

Fig.20 Stream Graph for Raw frequency words in Corpus

The above figure shows Stream Graph for High frequency words with reference to the chapter wise division in corpus.

Textual Arc for corpus

The Textual Arc in Voyant Tools is a circular visualization showing the distribution and connections of words in the corpus, blending text, frequency, and relationships for interactive exploration. It combines a circular layout with a word frequency visualization, where words are displayed along the arc, sized by frequency, and linked to show relationships or co-occurrences within the corpus.

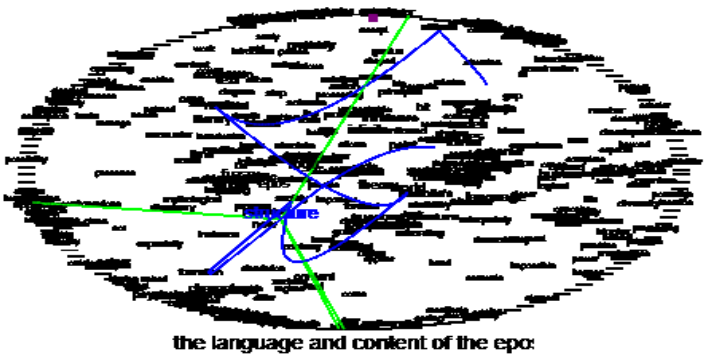

Fig. 21 Textual Arc in Corpus

The above figure shows Textual Arc with a word frequency visualization, where words are displayed along the arc, sized by frequency, and linked to show relationships or co-occurrences within the corpus which provides a comprehensive understanding of word frequency and the connections between terms with the corpus's indicated page numbers in a visually appealing manner.

Topics Tool for Corpus

The Topics tool in Voyant Tools identifies and visualizes clusters of words that frequently appear together, representing potential themes or topics within the corpus.

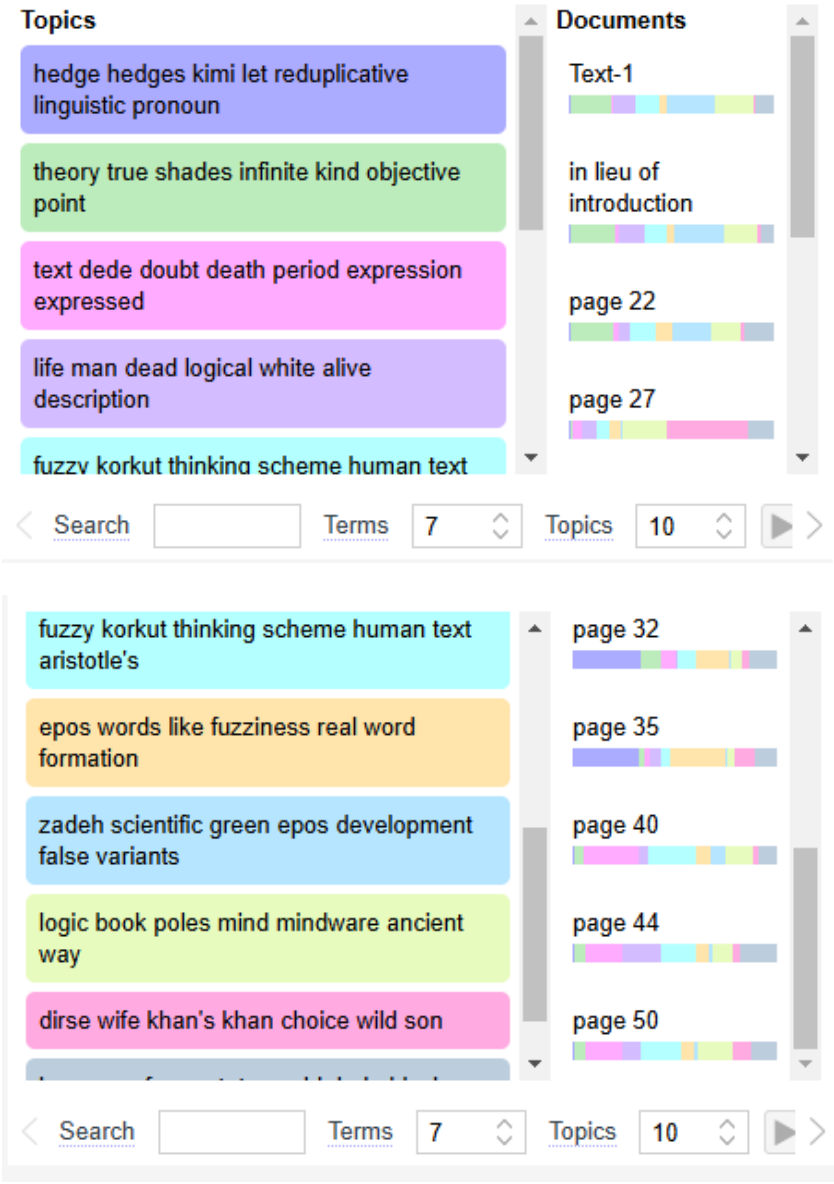

Fig. 22 Topics in Corpus

Bubbles in Voyant Tools

The Bubbles visualization in Voyant Tools shows the frequency of selected terms as expanding and contracting bubbles, providing an interactive way to observe their relative prominence over time or across documents. The following figure shows Bubbles with a word frequency visualization, where words are displayed along with contracting

Bubbles, and it helps in corpus analysis by visually comparing term frequencies and identifying trends and patterns, such as Logic, fuzzy, text, mind, language, like, thinking terms are most prominent or how their usage changes across sections in document.

Fig. 23 Bubbles in corpus

Terms Berry for Top Words

The Terms Berry in Voyant Tools displays a circular network of top words, with connections showing relationships or co-occurrences between them, aiding in understanding term associations within the corpus.

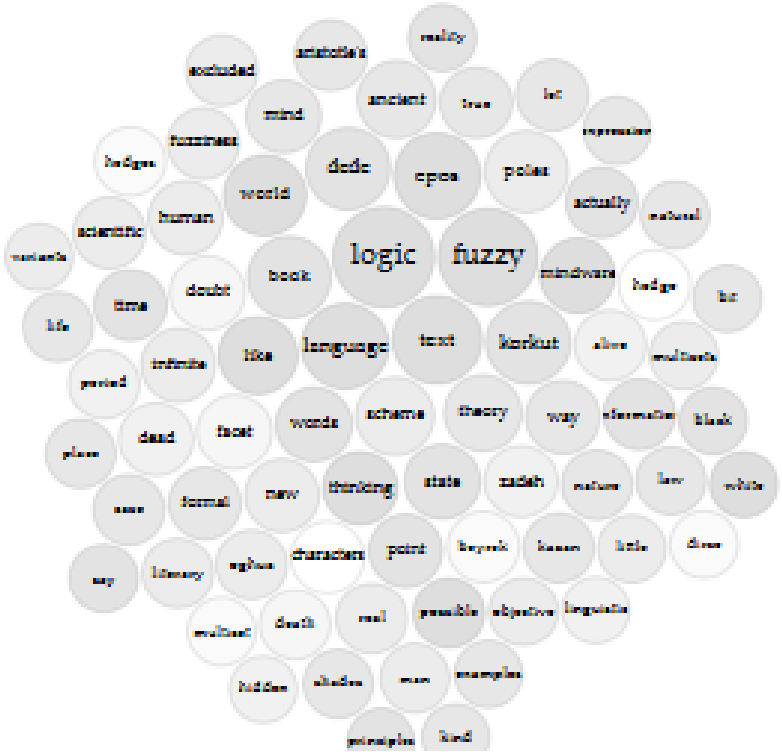

Fig. 24 Terms Berry for Top Words

The above figure of Terms Berry visualization helps by highlighting the relationships between the top words in the corpus as Logic, Fuzzy, Mindware, Language, Korkut, showing how frequently they co-occur or are connected. Moreover, it enables readers to identify significant themes and patterns in the text by providing them with information about word associations and relationships within context. The tool is highly interactive for users as they can click on single words in order to inspect their relationships and adjust settings in order to further refine their analysis.

Conclusion

In conclusion, the study has demonstrated advantages of applying a computational method to stylistic text analysis. Voyant Tools offer an easy and efficient means of investigating the Book of Dede Korkut corpus using fuzzy logic. (Abdulla and Aliev, 2023) see that elasticity and vagueness, which is characteristic of human language explained by fuzzy logic. With statistical tools, contextual examination, and visualization, its divergent options enabled both quantitative and qualitative observations, extending a deeper understanding of the literary and linguistic richness of the text under fuzzy logic. Data visualization, context investigation, and word frequency statistics were facilitated by Voyant Tools for text analysis. Voyant Tools provided a simple and effective way of probing the Book of Dede Korkut corpus based on fuzzy logic enabling computational approach to stylistic text analysis. Through statistical instruments, contextual analysis, and visualization, its multifaceted options provided quantitative and qualitative findings, and deepened comprehension of the literary and linguistic complexity of the text based on fuzzy logic on context research, and word frequency reports. The text's intricacy, cultural relevance, and stylistic depth provides a combination of fuzzy logic and digital analysis. By facilitating the identification of important language clusters, tone shifts, and narrative focus changes, Voyant Tools enhances the investigation of stylistic variation through Dream Scape and textual arc. Identification of stylistic variation across dialogue, description, and logical reasoning blocks in the text is facilitated by the ability to track co-occurrence patterns. The tool indicates whether "logic"-related words occur more frequently in analytical sections and if narrative parts exhibit other vocabulary options, thereby offering an indication of the stylistic entanglement of computational logic and conventional narratives.

Recommendations

Future research should consider integrating more advanced natural language processing tools such as TagAnt and Voyant Tools, and transformer-based models like BERT, and ALBERT to complement existing functionalities and allow for deeper stylistic patterns and linguistic analysis. Expanding the corpus to include a wider range of oral and literary texts from different cultures could help test the generalizability and robustness of the fuzzy logic framework. Enhancing the keyword-in-context (KWIC) analysis through the integration of sentiment detection or metaphor identification algorithms would also offer more nuanced interpretations of thematic and rhetorical elements. Developing customized visual dashboards or analytical plugins could improve the usability of Voyant Tools for researchers working on complex corpora. Finally, interdisciplinary collaboration among linguists, literary scholars, and computational scientists is recommended to refine the application of fuzzy logic in textual analysis and to explore new directions for computational approaches in the study of traditional narratives.

References

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Gregory, K., Geiger, L., & Salisbury, P. (2022). Voyant Tools and descriptive metadata: A case study in how automation can complement expertise knowledge. *Journal of Library Metadata, 22*(1–2), 1–16. https://doi.org/10.1080/19386389.2022.2030635

Gulati, S. (n.d.). Understanding the research landscape of smart libraries using text mining and data visualization: A use case of Voyant Tool.

Heuser, R., & Le-Khac, N. (2012). Learning to read data: Bringing out the humanistic in the digital humanities. *Victorian Studies, 54*(1), 79–86.

Hoover, D. L., Culpeper, J., & O'Halloran, K. (Eds.). (2014). *Digital literary studies: An overview*. Routledge.

Horenberg, L. (n.d.). Voyant Tools' little outing: How a text reading and analysis environment can help literary translators.

Liu, X. (2011). *Text mining and visualization: Case studies using open-source tools*. Wiley.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

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

Sampsel, L. J. (2018). Voyant Tool. *Music Reference Services Quarterly, 21*(3), 153–157. https://doi.org/10.1080/10588167.2018.1496754

Sinclair, S., & Rockwell, G. (2016a). *Embedding Voyant Tools*. Voyant Tools.

Sinclair, S., & Rockwell, G. (2016b). *Getting started*. Voyant Tools.

Sinclair, S., & Rockwell, G. (2016). *Voyant Tools: See through your text*. Retrieved from https://voyant-tools.org

Ullah, Z. (2022). *Unveiling knowledge pattern in intermediate English textbooks through Voyant text mining tools: A digital humanities study* (Unpublished doctoral dissertation). National University of Modern Languages, Islamabad, Pakistan.

Varshney, A. K., & Torra, V. (2023). Literature review of the recent trends and applications in various fuzzy rule-based systems. *International Journal of Fuzzy Systems, 25*, 2163–2186. https://doi.org/10.1007/s40815-023-01534-w

Wright, J. (2020). Researchers, teachers, and learners seeing new possibilities with Voyant Tool.*TESL-EJ,24*(94).