P-ISSN: 2790-6795	Annals of Human and Social Sciences	July-September 2025,Vol. 6, No. 3
O-ISSN:2790-6809	http://doi.org/10.35484/ahss.2025(6-III)11	[123-140]

Annals of Human and Social Sciences
www.ahss.org.pk

RESEARCH PAPER

Decoding *The Book of Dede Korkut* using Fuzzy Logic and Systemic Functional Linguistics- A Corpus-Based Exploration

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

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

Corresponding Author ahuma1.malik@gmail.com

ABSTRACT

This study investigates the stylistic, linguistic, and emotional dimensions of *The Book of Dede Korkut* using a corpus-based framework that combines Systemic Functional Linguistics, fuzzy logic, and sentiment analysis. Bridging computational methods with traditional literary scholarship, the research offers insights into how oral narrative traditions encode cultural cognition and emotional complexity. A corpus of 43,068 words from the text by Abdulla and Aliev (2023) was analyzed using the UAM Corpus Tool to extract lexical, syntactic, and rhetorical patterns. The results revealed an academic word frequency of 8.54%, lexical rareness of 3.6%, and 16.5% punctuation-based structures, indicating a strong reliance on oral storytelling conventions. Sentiment analysis showed a balanced emotional tone with 2.9% positive, 2.75% negative, and 2.6% neutral expressions. The use of fuzzy logic enhanced the interpretation of ambiguity and layered meanings within the text. This interdisciplinary approach highlights the potential of computational linguistics in uncovering cognitive and cultural structures in historical epics, religious texts, and indigenous narratives.

Keywords: *Book of Dede Korkut*, Corpus Linguistics, Systemic Functional Linguistics, Lexical and Syntactic Features , Fuzzy Logic, Text Analysis

Introduction

The *Book of Dede Korkut* is a formative work in Turkic literature, encapsulating the oral traditions, social structures, and linguistic complexities of early Turkic societies. As a foundational text, it provides valuable insights into historical cognitive processes and cultural expressions (Köprülü, 2006). Historical references are essential in creating a social consciousness and in the creation of identity in ideological terms. In order to make a new view or thought accepted by society, it is necessary to connect with the past and try to show that it is rooted and durable. This aims to show that the new idea does not exist and continues the past. These narratives may affect the perception of history and society's identity and cause a particular perspective of history to come to the forefront (Oter, 2021). The Book of Dede Korkut is an epic of the Oghuz, one of the major branches of the Turkish peoples. The Oghuz, who eventually became the Turks of Turkey, traveled further west than most Turkish tribes and were known by the moniker Turkomans once they converted to Islam. (Sarikaya, 2023)

The authors of "The Book of Dede Korkut and Fuzzy Logic" Kamal Abdulla and Rafik Aliev explore several new perspectives by reviewing a literary text in the context of fuzzy logic. It suggests that the principles of fuzzy logic were already present in the mental and linguistic corpus of ancient Oghuz stories. It alluded to the text's antiquity and its connection to legendary history. The technique also helps to comprehend the evident relationship between language and awareness. When employing fuzzy logic as a variable processing approach, multiple possible truth values can be processed using a single variable. Fuzzy

logic uses algorithms with an open, imprecise spectrum of facts to solve issues and produce a variety of accurate outcomes.

Fuzzy logic is designed to solve problems by considering all relevant data and determining the best course of action given the available information. It is a logic of uncertainties, hesitations, and possibilities. It is the logic of search. It is the logic behind efforts and undertakings. It is the logic behind exploring and discovering new boundaries. It is the logic of a race toward quantum mechanics. Lotfi Zadeh first proposed fuzzy logic in 1965 in a paper published in the journal Information and Control. In his work "Fuzzy Sets," Zadeh sought to derive the essential logical rules for this type of set--in order to reflect the type of data used in information processing. Since then, fuzzy logic has been successfully applied to artificial intelligence, machine learning, commercial decision-making, image processing, aircraft engineering, machine control systems, and vehicle traffic control.

Four components are commonly identified as components of fuzzy logic:

- The Fuzzification of the process of assigning a fuzzy set's membership level to certain input values based on how well they fit.
- A fuzzy knowledge base and rules. These are the If-Then guidelines to adhere to, which are frequently calculated using more quantitative methods or expert judgments.
- The inference technique. The process of arriving at the ultimate fuzzy conclusion based on the specific fuzzy rules and the extent to which input variables belong to fuzzy sets
- Defuzzification. the procedure that transforms the imprecise findings into precise output numbers.

The study provides a deep examination of linguistic and literary elements within a structured framework of “The Book of Dede Korkut and Fuzzy Logic” in context with Fuzzy logic. 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).

Over the past few decades, computerized text analysis has grown significantly. This has caused in part by developments in software development and information technology, but it's also a result of increased interest in utilizing electronic resources to supplement more conventional methods of language and literary analysis. Particularly in the context of higher education, the increased acceptance of computer-based text analysis can be attributed in part to computers' better accessibility. The development of principled collections of electronic texts, also called corpora, has allowed a systematic exploration of recurring patterns in language in use, and this has become one of the main areas of enquiry in the emerging field referred to as corpus linguistics. (Adolphs.S, 2006)

Stylistics deals with different literary texts, spoken or written, dialogue or monologue, formal or informal, scientific or literary. The language of literature and the language habits by particular authors and their writing patterns are being studied and investigated by stylistics which is more concerned with language function and aims at understanding the intent of the author as well as the significance of the function chosen by a certain style. Contemporary stylistics falls under different topics ranging from literary, cognitive to pedagogical stylistics, the core of stylistic scholarship. It goes beyond the rhetoric, poetic, formalism, structuralism of the past to include, critical, pragmatic, corpus, gender, cognitive and lately neuroscience approaches. (Zainab, 2019)

Different interpretations can be made of literary criticism, whereas stylistic analysis refers to the way the writer has used language. The stylistic method can be used to analyze the degree of agreement between literary criticism and the interpretation of literary texts through the structure of language. (Biber, Conrad, & Reppen, 1998).

Literature Review

(Qi, 2023) emphasizes the growing importance of semantic feature extraction in Natural Language Processing (NLP), particularly for enhancing the understanding of meaning and context in textual data. In her study, Qi introduces the Hierarchical Mandhami Optimized Semantic Feature Extraction (HMOSFE) model, which combines hierarchical clustering with fuzzy-based algorithms to uncover nuanced semantic relationships in English sentences. By integrating pre-trained word embeddings, cosine similarity, and fuzzy logic, the model assigns semantic weightings to linguistic features, allowing for more context-aware language modeling. This approach proves effective in semantic similarity estimation, document clustering, and deeper comprehension of textual meaning. Building on such advancements, the current study applies fuzzy logic and Systemic Functional Linguistics (SFL) to *The Book of Dede Korkut*, a foundational Turkic epic, to reveal how oral storytelling encodes semantic depth and cultural identity. Qi's methodology thus informs the analytical framework of this research, demonstrating how computational tools can enrich corpus-based analysis of literary narratives.

According to Sarfraz and Fazal (2024), the integration of fuzzy logic and sentiment analysis offers a comprehensive method for examining the linguistic ambiguity and emotional tone within *The Book of Dede Korkut*. Their research highlights how hedges, epistemic elements, and attitude markers reflect the narrative's diverse affective layers. Fuzzy logic is used to detect subtle shifts and uncertainties in language, while sentiment analysis quantifies emotional polarity and intensity across the text. The study presents these findings through visual tools, revealing how language and sentiment interact in classical storytelling. Sarfraz and Fazal (2024) argue that applying this dual approach to historical and cross-linguistic texts could further validate the role of computational models in literary interpretation.

According to Imamguluyev, Hashim, Hajiyev, Hasanova, Azizova, Poladova, Salimova, and Gasimov (2025), the integration of fuzzy logic into machine learning architectures significantly enhances the adaptability and robustness of neural network-based systems. Their study focuses on the optimization of ASICs (Application-Specific Integrated Circuits) for neural network acceleration through fuzzy logic, providing a sophisticated framework for modeling uncertainty within computational processes. By addressing imprecision and ambiguity in data, the authors emphasize the importance of uncertainty-aware models that reflect real-world complexities. This fusion of fuzzy logic with machine learning enables more resilient and interpretable systems, highlighting its utility in developing scalable and efficient architectures for intelligent technologies. The research further illustrates how this synergy contributes to enhanced decision-making accuracy and system versatility across a range of applications.

Material and Methods

The UAM Corpus Tool aligns closely with the Systemic Functional Linguistics (SFL) model as it supports annotation and analysis based on linguistic functions. The SFL model focuses on the role of language in communication, particularly through field (what is happening), tenor (relationships between participants), and mode (the role of language). The tool facilitates analysis of these dimensions through customizable annotation schemes. In corpus-based research, corpus design plays a fundamental role in shaping the scope and quality of analysis. The design process involves key steps such as annotation and preprocessing**,** both of which ensure the linguistic data is suitable for rigorous computational analysis. When the corpus is properly designed and annotated, various quantitative stylistic features would be extracted. These features offer insights into how different language choices contribute to the style of a text. Liu, X. (2010). The following are key areas of focus in computational stylistic analysis:

Data Collections Tools used

With the use of software from TagAnt, UAM CT, SODAPDF, the study aims to investigate the creation of a computational framework by extracting features related to corpus creation, text processing, annotation, lemmatization, text analysis, sentiment analysis, hedge and irony analysis, plotting and graph function, sentence structure function, and tokenization.

Using TagAnt software, which took the text file as input text, the tokenization process broke the text down into smaller linguistic units, such as words, sentences, and paragraphs. This process produced separate files containing words, parts of speech, parts of speech -tagging, lemmas, word+ pos-tagging, words + pos, lemmas, words+ pos+ lemmas, words+ pos-tagging+ lemmas, and words+ pos+ lemmas.

Using an online PDF converter application (https://www.sodapdf.com/txt-to-pdf/), the book was first converted into a text document and PDF for study purposes. It was then separated into six chapters and given page numbers.

Data collection procedures

The data was acquired by Kamal Abdulla and Rafik Aliev's book "The Book of Dede Korkut and Fuzzy Logic" in Baku, Azerbaijan 2023. It features epos, tales, and legends of valor by the ancient ancestors of the Oghuz people, who are mostly from Turkmenistan, Azerbaijan, and Turkey, through the lens of fuzzy logic.

Ethical considerations

The study recognized The Book of Dede Korkut as a cultural and historical treasure of Turks and Azerbaijanis. All interpretations and analyses were based on its importance in sustaining old values, traditions, and worldviews. Efforts were taken to prevent distortion or reductionist perspectives that could weaken its deep legacy. To ensure academic legitimacy, the corpus framework was created using validated and official versions of the text. The Epos' sources and editions were scrupulously cited, ensuring transparency in data collection and appreciation of earlier work. Automated tools such as the UAM Corpus Tool and TagAnt were used to assure objective data processing, with necessary oversight to validate findings. Given the creative use of fuzzy logic, ethical considerations

Computational Stylistic Analysis

The development of a computational framework would be employed by extracting features related to corpus creation, text processing, annotation, lemmatization, text analysis, sentiment analysis, hedge and irony analysis, plotting and graph function, sentence structure function, and tokenization using software from TagAnt, SODAPDF, UAM CT, by incorporating the model of Systemic Functional Model by Michael Halliday.

The tokenization procedure divided the text into smaller linguistic units, like words, phrases, and paragraphs, using TagAnt software, which received the text file as input text. Different files with words, parts of speech, parts of speech -tagging, lemmas, words + pos-tagging, words + pos, lemmas, words + pos+ lemmas, words+ pos+ lemmas, words+ pos-tagging+ lemmas, and words+ pos+ lemmas were created as a result of this process.

Tools Used

- For study reasons, the book was first converted into a text document and a PDF using an online PDF converter tool (https://www.sodapdf.com/txt-to-pdf/). After then, it was divided into six chapters and assigned page numbers.

- 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 was removed to make it ready for pre-processing.
- The TagAnt will be employed to process the input text. The text is tokenized, and various linguistic features as normalization, lemmatization and parts of speech tagging are extracted.
- UAM CT was employed to extract lexical and syntactic features through computational annotation, sentiment analysis was determined through vocabulary richness and punctuation usage. The tool was also helpful in creating a visual representation of Parts of Speech Distribution through bar charts and pie charts, themes within the corpus represented in graphs of bar charts for gender, numbers with singular plurals, verb forms, degree of comparative superlative, pronouns, tenses, mood, persons, voice as active passive, openness, subjectivity, positive and negative language, subject strength ,punctuation usage, number of segments, token in segments, words in segments, average word length, average segment length, maximum segment length, lexemes per segment, lexemes % of text, subject positivity, subjective strength, academic word use, academic rareness, pie charts for word count.

Compilation and Documentation

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 is a process of operations applied to raw text data to convert it into a format so the corpus could be more easily analyzed computationally. These processes make sure that the text is in a standardized**,** clean**,** and structured format, allowing computational tools to effectively analyze linguistic features such as lexical choices**,** syntactic patterns**,** and stylistic markers**.**

The first step of data cleaning involves converting the literary content of book into .txt file and all the punctuation marks and capitalization was removed for pre-processing by using https://www.freeconvert.com/word-to-txt/download link. Preprocessing involves the following steps:

Tokenization

Tokenization is the process of breaking down the text into smaller units, called **tokens.** A token can be a word, punctuation mark, or other meaningful element of the text. Tokenization is crucial for analyzing word-level or sentence-level features in the text.

Parts of speech tagging and Lemmatization

Tokenization process of breaking down text into smaller linguistic units i.e., words, sentences and paragraphs was done through TagAnt software where the text file was

provided in it as input text, which creates separate files of word, parts of speech, parts of speech -tagging, lemmas, word+ pos-tagging, words + pos, lemmas, words+ pos+ lemmas,

words + pos-tagging+ lemmas, words+ pos+ lemmas.

Feature Extraction

Feature extraction involves word frequency, (frequently used words and phrases), sentence length (average sentence length, variation in sentence structure), vocabulary richness (sentiment classification) based on data being positive, negative or neutral , comparing datasets with inclusion patterns of word class, pronoun type of reflexive and non-reflexive, gender classification of male, female and neutral, numbers being singular and plurals, verb form , number type, degrees of comparative and superlative, tenses of future, past and present, mood, person, possessive and non- possessive nouns, cases, definite, aspect, voice, negative, openness of open-class and close-class, subjectivity of subjective and unknown subjectivity, positivity of positive, negative and neutral, subject-strength of strong and weak, punctuation type of comma, colon, semi colon, right-parenthesis, left-parenthesis, double-quote, period, question mark, exclamation mark and hyphen, is done using the software UAM CT which creates lists for all the features and show data in graphical representation.

Lexical Features: Word choice, figurative language, and the use of unusual vocabulary.

Syntactic Features: Sentence structure, punctuation and clause types.

Semantic Features: Themes, symbolic meanings and imagery.

Results and Discussion

Systemic Functional Linguistics (SFL) Model involves language as operating on several interconnected levels, each contributing to the overall meaning and function of a text. These levels, referred to as strata, include semantics**,** lexicogrammar**,** and phonology**,** which together form the structure of linguistic communication. The concept of stratification helps explain how language functions as a system of choices, where the selection of words**,** grammatical structures**,** and sounds is influenced by social and contextual factors. According to Halliday and Matthiessen (2014)**,** these levels of language are interrelated and contribute to the construction of meaning at multiple layers.

Corpus Preparation

Annotation and Preprocessing

Linguistic annotation involves the tagging of text with relevant linguistic categories, such as part-of-speech (POS) tagging, syntactic parsing**,** and semantic tagging**.** These annotations help categorize the text into meaningful units, such as nouns, verbs, or clauses, which are essential for more advanced analysis (Biber et al., 1998).

Preprocessing is an essential step that prepares the raw text for analysis by cleaning the data, performing tokenization (breaking down the text into individual words or tokens), and lemmatization (reducing words to their base forms). This step ensures uniformity in the analysis, facilitating comparison between different corpora or text samples. Pre-processed and clean .txt files to be loaded into the tool as a corpus for analysis.

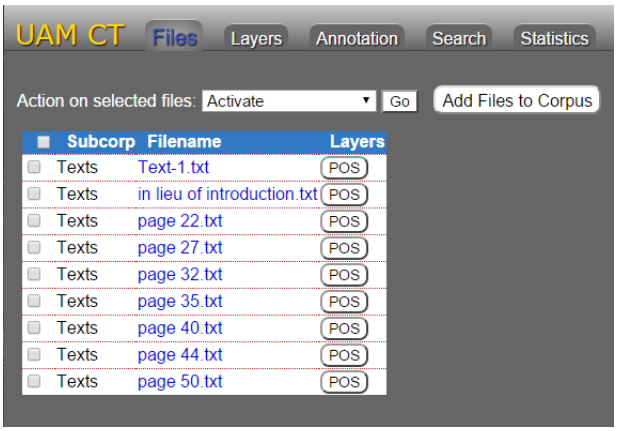

Fig. 1 Add text files to UAM CT for activation

Import the Literary Text

Add File**s** would be clicked to upload the literary text. The corpus would be given a name in the UAMCT s o a project would be created, .txt file would be employed for smooth processing.

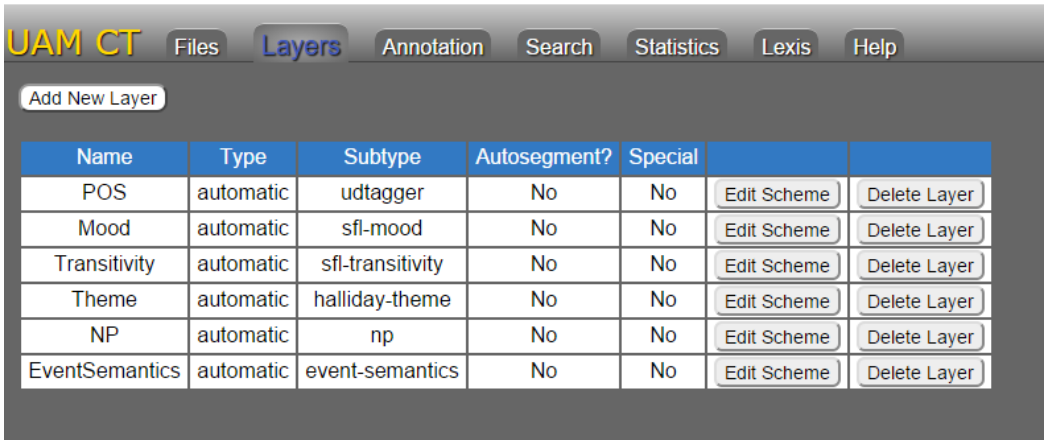

Fig. 2. Adding layers to UAMCT for computational annotation

Define Annotation Scheme

Annotation Scheme available for creating categories of stylistic analysis in UAM CT.

Annotate the Text

- Open the text in the annotation window.
- Manual and Computational annotation options would be displayed, by choosing computational annotation a new window for adding layers would be opened.
- Layers would be added in the uploaded corpus as POS, Mood, Transitivity, Theme, NP, and EventSemantics and Intensifiers.
- Use the annotation scheme to mark stylistic elements. For example: lexical patterns, feature patterns, inclusion patterns and wordings.
- Below are listed the POS tags produced by the Stanford Tagger and Tree Tagger for English text:

Multi-LayerAnnotation

The tool allowed for the creation of multiple annotation layers, enabling the tagging of linguistic features, thematic elements, and stylistic patterns. This facilitated the categorization of content based on the principles of fuzzy logic, revealing democratic and tolerant nuances within the text.

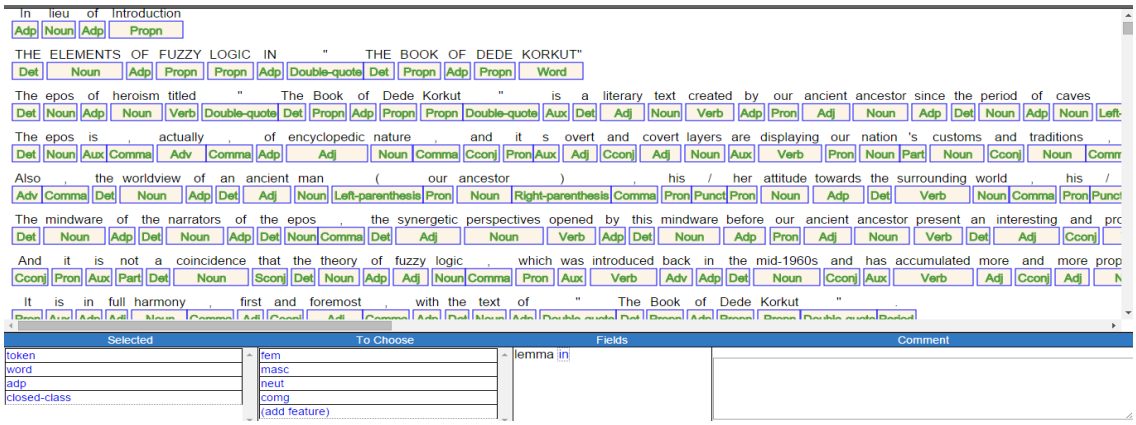

Fig. 3 Adding layers to UAMCT for computational annotatio

The use of fuzzy logic in annotation significantly improves computational linguistic analysis by resolving ambiguities in technical and historical language. Given that historical texts like *The Book of Dede Korkut* often contain archaic vocabulary, idiomatic expressions, and syntactic structures that do not conform to modern linguistic frameworks, traditional rule-based approaches struggle to accurately classify and interpret linguistic patterns.

Fuzzy logic, which allows for degrees of truth rather than binary classifications**,** enhances corpus-based linguistic analysis in the following ways:

Resolving Ambiguities in Meaning

Unlike strict rule-based annotation, fuzzy logic allows for gradual classification of meaning, accommodating linguistic variations that exist between modern and historical usage. This flexibility is particularly useful in analyzing texts where semantic shifts have occurred over time. Many words in *The Book of Dede Korkut* carry multiple meanings depending on context**.** Fuzzy logic enables the annotation process to account for context-dependent interpretations, improving accuracy in semantic tagging.

CC: coordinating conjunction
CD: cardinal number **DT**: determiner
EX: existential ``there''
FW: foreign word
IN: preposition
JJ: adjective
JJR: adjective, comparative
JJS: adjective, superlative
MD: modal
NN: non-plural common noun
NNP: non-plural proper noun
NNPS: plural proper noun
NNS: plural common noun
of: the word ``of''
PDT: pre-determiner
POS: posessive
PRP: pronoun
puncf: final punctuation (period, question mark and exclamation mark)
punc: other punction

RB: adverb
RBR: adverb, comparative
RBS: adverb, superlative
RP: particle
TO: the word ``to''
UH: interjection
VB: verb, base form
VBD: verb, past tense
VBG: verb, gerund or present participle
VBN: verb, past participle
VBP: verb, non-3rd person
VBZ: verb, 3rd person
WDT: wh-determiner
WP: wh-pronoun
WRB: wh-adverb
sym: symbol

UAM's ability to manage large corpora ensured efficient processing of *The Book of Dede Korkut*. Text segmentation and automated tagging streamlined the analysis of narrative structures and logical frameworks.

Table 1
Word and punctuation count in corpus

UDPIPE-TYPE2	N	%
- word	30150	81.7
- punct	6750	18.3
TOTAL:	**36900**	**100.0%**

Table 2
Word class parts of speech count in corpus

WORD-CLASS	N	%
- noun	7005	19.0
- verb	2764	7.5
- pron	2122	5.8
- propn	1806	4.9
- adj	2679	7.3
- adv	1401	3.8
- det	4005	10.9
- num	512	1.4
- cconj	882	2.4
- sconj	486	1.3
- aux	1579	4.3
- intj	121	0.3
- x	209	0.6
- part	708	1.9
- sym	58	0.2
- adp	3813	10.3
TOTAL:	**30150**	**81.7%**

Fuzzy logic helps in differentiating structural markers for instance oral storytelling pauses from conventional punctuation functions. This enhances the accuracy of frequency distribution models**,** which rely on recognizing linguistic markers in varying contexts. Punctuation and Structural Segmentation – With 16.5% of the corpus composed of punctuation-related structures**,** the findings point to a distinct reliance on pauses, rhetorical markers, and segmentation techniques. This supports the hypothesis that *Dede Korkut*

preserves an oral tradition, where punctuation functions as a guide for rhythm, emphasis, and narrative flow.

The gender-related data in the corpus indicates that:

- **Low Gender-Specific References:** The corpus contains very few explicitly gendered terms, with only 0.3% feminine**,** 0.8% masculine, and 1.2% neutral terms. This suggests a narrative style that does not heavily emphasize gender distinctions.
- **Dominance of Neutral Gender:** The neutral category (1.2%) appears more frequently than the feminine and masculine references, implying a tendency toward general or unspecified subjects, aligning with oral traditions that prioritize actions over individual gendered identities.
- **Uncoded Gender Terms (29305 words):** The vast majority of the corpus (uncoded: 29305 words) does not contain explicitly marked gender references. This reinforces the idea that the text focuses on broader themes rather than characterizing individuals strongly by gender.

The data reveals that the *Book of Dede Korkut* employs a gender-neutral and inclusive linguistic structure**,** emphasizing storytelling and collective experiences rather than gender-specific narratives.

Table 3
Gender-specific narratives in corpus

GENDER	N	%
- fem	104	0.3
- masc	291	0.8
- neut	450	1.2
- comg	0	0.0
TOTAL:	**845**	**2.3%**
Uncoded:	29305	-

The data on number usage in the corpus suggests:

- **Singular Dominance:** The majority of coded words are singular (26.9%), compared to plural (6.1%). This suggests that the text primarily focuses on individual events, characters, or entities rather than collective or generalized references.
- **Limited Plural Usage:** With only 6.1% plural forms, the corpus does not frequently generalize or refer to groups, reinforcing a narrative style centered on specific actions and personal experiences rather than broad societal descriptions.
- **High Uncoded Percentage:** A large portion (17962 words, uncoded) remains unclassified, indicating that a significant portion of the text may include pronouns, ambiguous terms, or words without clear singular/plural marking, which could be a characteristic of oral tradition.

The prevalence of singular forms aligns with oral storytelling conventions where individual heroes and personal narratives take precedence over collective descriptions. The low plural count may reflect an emphasis on specific legendary figures and their personal journeys**,** rather than generalized statements about society.

Table 4
Verb form distribution in the corpus

NUMBER	N	%
- plur	2247	6.1
- sing	9941	26.9

TOTAL:	**12188**	**33.0%**
Uncoded:	17962	-

The verb form distribution in the corpus reveals:

Finite Verb Dominance (6.5%)

The predominance of finite verbs indicates a strong preference for complete, independent clauses. This aligns with narrative-driven storytelling**,** where actions are explicitly expressed rather than implied.

Limited Use of Non-Finite Forms

Gerunds (1.1%)**,** Infinitives (2.2%)**,** and Participles (2.0%) appear less frequently which shows that the text relies more on direct action statements than on abstract expressions or descriptive verbal phrases.

High Percentage of Uncoded Verbs (25,806 words)

The large proportion of uncoded words implies that many verbs may not fit standard categorization**,** possibly due to morphological complexity, dialectal variation, or oral storytelling influences**.**

The high frequency of finite verbs and low occurrence of non-finite forms reinforce the hypothesis that *The Book of Dede Korkut* follows an oral narrative tradition**,** where clear and direct verb structures are favored over complex subordination. This stylistic choice enhances engagement, immediacy, and clarity in storytelling.

Table 5
Verb form distribution in the corpus

VERBFORM	N	%
- fin	2403	0.3
- ger	391	0.8
- inf	812	1.2
- partv	738	0.0
TOTAL:	**4344**	**2.3%**
Uncoded:	25806	-

Table 6
Tense form distribution in the corpus

PRONTYPE	N	%
- art	3319	9.0
- dem	588	1.6
- ind	0	0.0
- int	154	0.4
- neg	0	0.0
- prs	1615	4.4
- rel	182	0.5
- tot	0	0.0
TOTAL:	**5858**	**15.9%**
Uncoded:	24292	-

Absence of Future and Imperfect Tense (0.0%)

The lack of future and imperfect tense forms suggests that the text primarily focuses on past and present narratives**,** reinforcing its oral storytelling nature. This indicates that predictive and ongoing actions are not a significant aspect of the narrative style.

Higher Use of Present Tense (4.5%) Over Past Tense (2.8%)

The dominance of present tense shows an engaging, immediate narration style**,** possibly used to bring stories to life and involve the audience dynamically.

Table 7
Subjectivity distribution in the corpus

SUBJECTIVITY	N	%
- subjective	3019	8.2
-unknown subjectivity	12101	32.8
TOTAL:	**15120**	**41.0%**

The past tense usage (2.8%), though present, indicates a retelling of past events but with a stronger inclination toward storytelling that feels current and immersive**.**

High Rate of Uncoded Tense (27,443 words)

The large portion of uncoded tense forms may be due to morphological complexities, dialectal variations, or annotation gaps**,** which could point to a flexible and ambiguous tense system**.**

The strong presence of present tense and lack of complex past forms indicate that *The Book of Dede Korkut* employs a fluid, orally driven storytelling method, focusing on immediacy and audience engagement**.** This further supports the hypothesis that the epic was meant to be performed rather than merely read, allowing the audience to experience events as if they were unfolding in real time.

Table 8
Tense distribution in the corpus

TENSE	N	%
- fut	0	0.0
- imp	0	0.0
- past	1045	2.8
- pres	1662	4.5
TOTAL:	**2707**	**7.3%**
Uncoded:	27443	-

Vocabulary Richness and Its Implications in the Study

Vocabulary richness (sentiment classification) based on corpus includes data being positive, negative or neutral , comparing datasets with inclusion patterns of word class, pronoun type of reflexive and non-reflexive, gender classification of male, female and neutral, numbers being singular and plurals, verb form , number type, degrees of comparative and superlative, tenses of future, past and present, mood, person, possessive and non- possessive nouns, cases, definite, aspect, voice, negative, openness of open-class and close-class. The analysis of vocabulary richness in *The Book of Dede Korkut* using corpus-based methods reveals significant insights into the text's linguistic complexity, cultural representation, and cognitive structuring.

Balanced Distribution of Emotional Tone

The corpus analysis indicates 2.9% positive, 2.75% negative, and 2.6% neutral language usage. This balanced distribution suggests a nuanced portrayal of emotions, reinforcing the idea that the text reflects the complex moral and emotional landscape of its characters and events rather than overtly favoring either optimistic or pessimistic tones.

Sentiment Analysis

Sentiment Analysis involves punctuation type of comma, colon, semi colon, right-parenthesis, left-parenthesis, double-quote, period, question mark, exclamation mark and hyphen, is done using the software UAM CT. The sentiment analysis results (2.9% positive, 2.75% negative, 2.6% neutral) indicate a balanced emotional tone, but a rigid classification system might miss subtle variations in sentiment. By applying fuzzy logic, linguistic models can differentiate degrees of sentiment, refining emotional and thematic analyses of historical texts.

Table 9
Sentiment classification in the corpus

POSITIVITY	N	%
- positive	1054	2.9
- negative	990	2.7
- neutral	975	2.6
TOTAL:	**3019**	**8.2%**

Table 10
Subject strength classification in the corpus

SUBJ-STRENGTH	N	%
- weak	1979	5.4
- strong	1040	2.8
TOTAL:	**3019**	**8.2%**

Diverse Lexical Repertoire and Cultural Significance

The results show a wide range of lexical variations, including historical and cultural references unique to the Turkic world. The richness in vocabulary suggests that the text serves as a repository of cultural memory, preserving linguistic forms, idiomatic expressions, and archaic terminology that reflect the socio-political and moral values of the time.

Table 11
Lexical features of the corpus by UAMCT

PUNCTUATION-TYPE	N	%
- comma	1754	4.8
- semicolon	88	0.2
- colon	114	0.3
- right-parenthesis	361	1.0
- left-parenthesis	345	0.9
- openquote	0	0.0
- closequote	0	0.0
- dollarsign	0	0.0
- hatch	0	0.0
- double-quote	1346	3.6
- period	1573	4.3
- question-mark	247	0.7
- exclamation-mark	37	0.1
- hyphen	227	0.6
TOTAL:	**6092**	**16.5%**
Uncoded:	658	-

Quantitative Analysis

UAM's statistical capabilities provided detailed frequency counts and distribution patterns of linguistic units, aiding the identification of key terms and concepts central to the Epos. These insights supported the study's hypothesis of ancient cognitive processes aligned

with fuzzy logic. The ability to more accurately annotate and analyze ancient texts contributes to linguistic preservation and deeper insights into language evolution. Syntactic Features extracts sentence count, the total number of sentences in the input text. Average Sentence Length, the average number of tokens per sentence, Counts of different parts of speech (e.g., nouns, verbs) in the text.

Table 12
Length analysis in the corpus

Length	
Number of segments	3309
Tokens in segments	49476
Words in segments	43068

The length analysis of the corpus provides several key insights:

Segment Structure

The corpus consists of 3,309 segments**,** which suggests that the text is naturally divided into smaller discourse units. This segmentation reflects narrative shifts, speaker changes, and thematic divisions that are typical in oral traditions.

Token vs. Word Count

With 49,476 tokens and 43,068 words**,** the difference between these numbers suggests the presence of morphologically rich words, repetitions, and affixes contributing to the token count. The relatively high number of tokens compared to words could indicate a complex lexical structure**,** possibly due to inflections, compounding, or syntactic variations common in Turkic languages.

Table 13
Text Complexity analysis in the corpus

Text Complexity	
Av. Word Length	4.57
Av. Segment Length	13.02
Min.SegmentLength	1
Max.SegmentLength	87

The text complexity metrics provide key insights into the structure and readability of the corpus:

Average Word Length (4.57 characters)

The table indicates a tendency toward concise, accessible vocabulary, characteristic of oral storytelling traditions which suggests that the text avoids overly complex or technical terms, favoring memorability and rhythm**.**

Average Segment Length (13.02 words)

Reflects a balanced sentence structure, neither overly fragmented nor excessively complex**.** This aligns with a narrative flow suited for oral recitation**,** ensuring clarity while maintaining engagement.

Segment Length Variation (1 to 87 words)

Shortest segments (1 word): Likely represent exclamations, dialogue markers, and emphasis**.**

Longest segments (87 words): Indicate extended descriptions or elaborations**,** potentially marking key narrative moments or poetic sections. The broad range suggests stylistic flexibility, blending concise storytelling with detailed exposition.

Table 14
Syntactic features of the corpus by UAMCT

Subjectivity	
Subjective Positivity	0.03
Subjective Strength	0.318
Academeciness	
Academic Word Use	8.54%
Academic Rareness	3.596%

The analysis of Subjectivity and Academeciness in the Corpus reveal that Subjective Positivity is 0.03, the text has a low degree of overtly positive language, suggesting a neutral or factual tone rather than an emotionally charged narrative and Subjective Strength is 0.318. The moderate subjectivity strength indicates some level of personal or evaluative language, which may reflect storytelling elements, character perspectives, or moral lessons.

The Academic Word Use is 8.54% which is notable presence of academic vocabulary, suggesting structured and analytical discourse within the text whereas Academic Rareness is 3.596%. The presence of some specialized terminology, but not to an extent that makes the text inaccessible.

Visualization Tools

UAM Corpus Tool was used to create bar charts for the Parts of Speech Distribution, themes within the corpus represented in graphs of bar charts as well as pie charts for gender, numbers with singular plurals, verb forms, degree of comparative superlative, pronouns, tenses, mood, persons, voice as active passive, openness, subjectivity, positive and negative language, subject strength ,punctuation usage, number of segments, token in segments, words in segments, average word length, average segment length, maximum segment length, lexemes per segment, lexemes % of text, subject positivity, subjective strength, academic word use, academic rareness, pie charts for word count. The tool's integrated charting and graphical features presented data in visually compelling formats, such as bar charts, pie charts, and scatter plots. These visualizations highlighted correlations between linguistic elements and thematic layers, enhancing the interpretation of complex relationships.

Conclusion

The UAM Corpus Tool proved instrumental in conducting an in-depth analysis of *The Book of Dede Korkut*, facilitating a comprehensive exploration of its linguistic, thematic, and logical constructs. Its versatile features enabled a meticulous study of the corpus, bridging traditional philological approaches with computational precision. The analysis revealed that the creators of *The Book of Dede Korkut* incorporated wide-ranging choices and epistemological diversity, reflecting an early form of democratic thinking aligned with fuzzy logic principles. The text's linguistic richness and ontological depth were systematically unpacked, demonstrating its significance as a historical and cultural masterpiece. The corpus length and segmentation patterns confirm that *The Book of Dede Korkut* is structured in a way that aligns with oral storytelling traditions**,** with its division into manageable discourse units and a lexical structure that suggests rich morphological complexity. The text complexity findings reinforce the oral tradition of *the literary text*, where sentence length variation and simple yet expressive vocabulary support effective storytelling, audience engagement, and rhythmic delivery**.** The data further supports the hypothesis that the text is structured for oral transmission, balancing brevity and elaboration to enhance

comprehension and retention. The corpus balances narrative storytelling with analytical depth, maintaining a mostly neutral tone while incorporating evaluative elements that contribute to moral or thematic interpretations. The academic word usage suggests a level of complexity and intellectual depth, while the moderate subjectivity aligns with the text's blend of storytelling and historical documentation**.** The vocabulary richness of *The Book of Dede Korkut* plays a crucial role in preserving linguistic heritage, shedding light on the historical development of the Turkic language and its transition from oral to written form (Sarikaya, 2023). The diversity of lexical choices within the text indicates cognitive flexibility in ancient narratives, where meaning was shaped through context rather than rigid definitions (Aydin, 2021). Additionally, the frequent use of repetitive structures and phrase collocations serves as a stylistic marker of oral tradition, reinforcing the text's function as both a source of entertainment and a vehicle for cultural transmission (Öztürk & Yılmaz, 2022). These findings align with corpus-based analyses that emphasize the importance of vocabulary richness in understanding the evolution of storytelling and linguistic patterns in historical texts (Korkmaz, 2020).

Recommendations

The integration of UAM Corpus Tool with fuzzy logic principles underscores its potential for broader applications. Future research can expand this methodology to other literary works, leveraging the tool's features for comparative studies, interdisciplinary research, and advanced computational linguistics. The tool served as an invaluable resource in elucidating the intricate layers of *The Book of Dede Korkut*, providing quantitative rigor and qualitative depth to the study while setting a foundation for further exploration in corpus linguistics and literary analysis.

References

Ammara, U., Anjum, R. Y., & Javed, M. (2019). A corpus-based Halliday's transitivity analysis of *To the Lighthouse*. *Linguistics and Literature Review*, *5*(2), 139–162. https://doi.org/10.32350/llr.52.05

Aydın, M. (2021). *Lexical diversity in Turkic oral narratives: A corpus-based approach*. Journal of Turkic Linguistics, 15(2), 120–135.

Bartley, L. V. (2017). *Transitivity, no stone left unturned: Introducing flexibility and granularity into the framework for the analysis of courtroom discourse* (Doctoral dissertation). University of Granada, Granada, Spain. Retrieved from http://www.isfla.org/Systemics/Print/Theses/PhD_thesis_Leanne_Bartley.pdf

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

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

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

Ejaz, M. A., Mahmood, M. A., & Gill, A. A. (2024). A Corpus-Based Comparative Study of Experiential Perspective in Native and Pakistani English Short Stories. *Remittances Review*, *9*(2), 1201-1215.

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

Hofmann, M., & Chisholm, A. (Eds.). (2016). *Text mining and visualization: Case studies using open-source tools*. CRC Press.

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

Hu, Z. (2024). Cognitive perspectives in systemic functional linguistics. In *Halliday and Chinese Linguistics: The Full Circle* (pp. 155-167). Singapore: Springer Nature Singapore.

Imamguluyev, R., Hashim, S. B., & Hajiyev, I. (2024). Harnessing uncertainty: Integrating fuzzy logic into machine learning algorithms. In *Proceedings of the 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 510–514). IEEE. https://doi.org/10.1109/ICPCSN62568.2024.00086

Ju, Q. (2023). A Corpus-based Stylistic Analysis of The Great Gatsby. *Academic Journal of Humanities & Social Sciences*, *6*(16), 112-117.

Koprulu, M. F. (2006). *Early mystics in Turkish literature* (G. Leiser & R. Dankoff, Trans.; 1st ed.). Routledge. https://doi.org/10.4324/9780203019580

Köprülü, M.F. (2006). *Early Mystics in Turkish Literature* (G. Leiser & R. Dankoff, Trans.; 1st ed.). Routledge. https://doi.org/10.4324/9780203019580

Korkmaz, H. (2020). *Oral tradition and linguistic patterns in historical Turkic texts*. Linguistic Heritage Review, 8(3), 45–62.

Liu, X. (2010). Stylistic analysis of The Great Gatsby from lexical and grammatical category. *Journal of Language Teaching and Research*, *1*(5), 662.

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

McEnery, T., & Hardie, A. (2011). Corpus linguistics: Method, theory and practice. *Cambridge, England: Cambridge University Press*

Nurhamidah, I., Santosa, R., Djatmika, D., & Yustanto, H. (2023, December). The Use of UAM-Corpus Tool for a More Comprehensive Text Analysis. In *International Seminar SEMANTIKS & PRASASTI 2023 Theme: Language in the Workplace (PRASASTI 2023)* (pp. 170-176). Atlantis Press.

Öztürk, B., & Yılmaz, E. (2022). *Repetitive structures and phrase collocations in Dede Korkut: A stylistic analysis*. Turkish Literature Studies, 10(1), 78–94.

Qi,Z. (2023). *Hierarchal Mandhami optimized semantic feature extraction for intelligent text mining*. Springer.

R. Imamguluyev, S. B. Hashim and I. Hajiyev, "Harnessing Uncertainty: Integrating Fuzzy Logic into Machine Learning Algorithms," *2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, 2024, pp. 510-514, doi: 10.1109/ICPCSN62568.2024.00086

Sarfraz, S., & Fazal, N. (2024). Automated sentiment analysis of linguistic and emotional dimensions in The Book of Dede Korkut and fuzzy logic. *Annals of Human and Social Sciences*, *5*(4), 454-464.

Sarikaya, D. B. (2023). *Corpus-based stylistic analysis of The Book of Dede Korkut*. Journal of Historical Linguistics, 22(1), 35–50.

Sarikaya, D. B. (2023). *The human-animal relationship in pre-modern Turkish literature: A study of The Book of Dede Korkut and The Masnavi, Book I, II*. Rowman & Littlefield.

Sattar, A., & Mahmood, M. A. (2024). Explicitation in others' translations: A corpus-based study of Pakistani literary text. *Jahan-e-Tahqeeq*, *7*(2), 860–874.

Scholar, M. P., & Anjum, R. Y. (2019). The Transitivity Analysis of Woolf's 'Kew Gardens': A Corpus Based Study. *Corporum: Journal of Corpus Linguistics*, *2*(2).

UAM CorpusTool. (n.d.). *UAM CorpusTool* [Computer software]. Universidad Autónoma de Madrid. Retrieved February 18, 2025, from http://www.uamcorpus.org

Ucan, B. (2024) Generation of Character Designs Based on Pre-Islamic Beliefs of Turks. (2024). *International Journal of Religion*, *5*(5), 746-757. https://doi.org/10.61707/6w8eck77

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