

Advancing Literary Analysis With Python: A Comprehensive Study of Simile Detection and Classification in the Translation of *Al-Abrat*

Rasha Talal Al Zahrawi¹ , Syed Nurulakla Syed Abdullah¹ , Akila Sarirete²,
and Muhammad Alif Redzuan Abdullah¹ 

Abstract

This study explores the intricate use of similes within *Al-'Abrat* ("The Tears") by Mustafa Lutfi al-Manfaluti, a cornerstone of modern Arabic literature. The research employs Abdul-Raof's classification framework, combining traditional Arabic rhetorical theory with computational methods, including Python-based algorithms, to detect and categorize similes. By analyzing 212 similes extracted from the English translation "The Tears" by Majid Khan Malik Saddiqui, the study highlights similes' cognitive and rhetorical functions in enriching emotional resonance and cultural depth. Key findings reveal distinct patterns in simile usage, particularly the prevalence of perceptible-perceptible and cognitive-cognitive categories, emphasizing the role of figurative language in fostering thematic and emotional engagement. The interdisciplinary approach bridges Arabic literary criticism, cognitive linguistics, and natural language processing (NLP), offering methodological innovations for simile detection and classification. While effective, the study acknowledges limitations, including reliance on a single corpus, subjectivity in manual validation, and the need for advanced machine learning models for nuanced analysis. This research contributes to enhancing the interpretive framework for Arabic rhetoric and broadening the application of computational tools in literary studies. Future directions propose expanding the corpus, integrating diverse rhetorical devices, and employing sophisticated NLP techniques to further uncover the richness of Arabic literary heritage.

Plain Language Summary

Exploring Similes In "The Tears" Using Python Algorithms

This study focuses on similes—a figure of speech that compares two things using words like "as" or "like"—in "The Tears" by Mustafa Lutfi al-Manfaluti, an important work in classic Arabic literature. Similes add depth and emotion to writing, helping readers better understand ideas through vivid imagery. The research combines traditional Arabic literary analysis with modern technology to uncover patterns in how similes are used.

By analyzing 212 similes from the English translation of "The Tears," the study explores how these comparisons contribute to the emotional and cultural richness of the text. A computer program written in Python was used to automatically find and categorize similes, making the process faster and more efficient. Similes were grouped into different types based on whether they compared tangible things, abstract ideas, or a mix of both.

¹University of Putra, Serdang, Malaysia

²Effat University, Jeddah, Saudi Arabia

Corresponding Author:

Rasha Talal Al Zahrawi, Building No. 3016, Al Narjis Street, Al Sahafah District, Secondary No. 7205, Riyadh 13321, Kingdom of Saudi Arabia.

Email: zahrawi.rasha@gmail.com



Key findings revealed that similes in “The Tears” often connect concrete images (like “the night”) with abstract ideas (like “hope”), showcasing the power of Arabic rhetoric to evoke deep emotional responses. While the computer program performed well, a manual review was still needed to ensure accuracy, highlighting the importance of human expertise in understanding literary texts.

This study also shows how technology can enhance the analysis of Arabic literature while respecting its cultural and linguistic uniqueness. By blending literary analysis with computational tools, the research provides new ways to study classic texts and suggests future improvements, like using more advanced artificial intelligence, to make the process even better.

Ultimately, the research highlights the richness of Arabic literary traditions and the potential of combining traditional and modern methods to deepen our understanding of language and storytelling.

Keywords

Python, transcreation, rhetoric, literary translation, machine learning

Introduction

Al-'Abrat by Mustafa Lutfi al-Manfaluti

Mustafa Lutfi al-Manfaluti (1876–1924), a key figure in modern Arabic literature, created *Al-'Abrat* (“The Tears”), a work renowned for its moral stories and essays that simplified complex Arabic genres (Al-Manfaluti, 1916). Though only translated once, this English version bridges linguistic and cultural gaps, introducing Al-Manfaluti’s (1924) poignant storytelling to a broader audience and highlighting the universal themes in his writing.

The Tears, English Translation by Majid Khan Malik Saddiqui

The Tears, translated by Majid Khan Malik Saddiqui in 1915, comprises 10 evocative tales of love, sacrifice, and hardship. Saddiqui’s unique background as a non-native speaker enriches the translation, illustrating the challenges and depth of intercultural translation. His efforts, guided by the Sufi scholar Hazarat Allama Pir Alaaddin Siddiqui, allow readers to experience the emotional core of Al-Manfaluti’s (n.d.) characters.

Themes in Al-'Abrat

The study investigates similes in Arabic rhetoric, particularly their structure, categorization, and cognitive implications, combining traditional literary analysis with modern computational methods. It focuses on themes in Al-Manfaluti’s *Al-'Abrat*, such as loss and social injustice, using Abdul-Raof’s (2006) classification of similes and integrating theories like conceptual metaphor (Lakoff & Johnson, 1980) to explore their emotional and cognitive effects.

Key objectives include analyzing similes in classical Arabic poetry and the Qurān, examining their influence

on reader perception, and applying natural language processing (NLP) to automate simile detection and categorization using Python algorithms. The research addresses gaps in understanding simile usage in Arabic rhetoric and contributes to literary criticism, cognitive linguistics, and NLP by providing a methodological framework for simile analysis. It highlights the potential of computational approaches to enhance Arabic text analysis while maintaining a strong foundation in linguistic and cultural understanding (Abdul-Raof, 2006; Al-Mubarak & Hill, 2023; Al-Sharafi, 2004; Yaseena et al., 2022; Zafer, 2007).

The interdisciplinary methodology incorporates insights from literary criticism, cognitive science, and computer science, offering theoretical and practical advancements in the study of similes within Arabic literature.

Hypothesis of the Study

This study is guided by the following hypotheses:

Hypothesis 1: The use of Python-based natural language processing techniques can effectively detect and classify similes in Arabic literary texts with a high degree of accuracy.

Hypothesis 2: The classification of similes based on Abdul-Raof’s framework will reveal distinct patterns that provide deeper insight into the rhetorical and cognitive functions of similes in classical Arabic literature.

These hypotheses will be tested through a combination of computational analysis and manual validation, ensuring a robust examination of the simile detection and classification process. The paper presents a structured exploration of similes within Arabic rhetoric (balāgha), integrating literary criticism with computational

linguistics. It begins with an introduction that sets the stage by highlighting the importance of similes in Arabic rhetoric and outlining the study's objectives, methodology, and theoretical basis. A thorough literature review follows, delving into existing research on Arabic rhetoric, particularly focusing on similes. Next, the study is grounded in Arabic rhetorical theory, cognitive linguistics, and computational approaches, as the theoretical framework is detailed. The methodology section describes the procedures for text selection, computational extraction, and categorization of similes, followed by a comprehensive analysis of the data. The findings are examined in the context of the theoretical framework, assessing their implications for the understanding of Arabic rhetoric and cognitive processing of similes. The paper concludes by summarizing key insights, acknowledging limitations, and suggesting avenues for future research, supplemented by a complete list of references and, if applicable, appendices providing Supplemental Material.

Literature Review

Arabic Rhetoric Features in Literary Texts: State-of-the-Art

Arabic Rhetoric. Rhetoric (*balāgha*) in Arabic refers to the strategic use of language to persuade and evoke emotions, embodying the artistic and persuasive expression of ideas (Moreau, 2003). Central to Arabic literary criticism, *balāgha* evaluates the artistic value of texts, analyzing elements such as similes, metaphors, and stylistic features like paronomasia. It is pivotal in interpreting classical Arabic poetry and the Qurān, reflecting its historical and literary significance (von Grunebaum, 2012).

Arabic rhetoric consists of three sub-disciplines, focusing on adapting language to resonate with audiences, analyzing sentence structures, and employing advanced rhetorical strategies (Balagha, 2020). It plays a crucial role in elucidating literary devices and enhancing the appreciation of Arabic literature's artistic depth, themes, and motifs (Rahman et al., 2013). Proficiency in *balāgha* also refines writing skills, fostering linguistic precision and persuasiveness, which are vital for writers and academics (Adeyemi, 2017).

However, challenges persist in distinguishing between rhetorical tropes like metaphor, simile, and hyperbole. While metaphor is often considered primary among figurative devices, debates about the classification of other tropes remain ongoing (Al-Mubarak & Hill, 2023; Al-Sharaf, 2004).

Arabic Rhetoric and Similes: Background and Significance

Arabic rhetoric boasts a profound historical foundation, originating in pre-Islamic times. Its growth and

formalization were heavily influenced by scholars' interest in studying the unique and inimitable (*i'jaz*) qualities of Qur'anic discourse (Abdul-Raof, 2006). Through centuries, Arabic rhetoric has developed into three principal branches: *'Ilm al-ma'ani* (focusing on word order and semantic syntax), *'Ilm al-bayan* (dedicated to figures of speech), and *'Ilm al-badi'* (concerned with rhetorical embellishments). Among these, the simile (*tashbih*) is highly regarded alongside metaphor as a crucial figure of speech (Alshammari, 2016). Recent research in this field has adopted diverse perspectives, including cultural and pragmatic analysis (Abdul-Raof, 2006), translation studies, and computational techniques (Hussein & Kuflik, 2023). These areas of study delve into the complexities of interpreting Arabic rhetorical elements within cultural contexts, translating similes across languages, and even employing technology to analyze classical texts. Moreover, innovative pedagogical methods are enhancing how Arabic rhetoric, particularly Qur'anic rhetoric (Grafiati, 2021), is taught, and comparative studies shed light on how Arabic rhetorical traditions align with or differ from those of other cultures (Kızıklı, 2007). This research enriches our understanding of Arabic rhetoric's adaptability, relevance, and depth across various academic fields.

Literature on Figurative Language and Similes in Literary Analysis

Figurative language, particularly similes, plays a crucial role in literary analysis and cognitive understanding of texts by helping readers interpret and connect with complex ideas through vivid imagery and emotional resonance (Admin, 2019). Figurative language includes devices like similes, metaphors, personification, symbolism, and hyperbole, all of which enable authors to evoke emotions and deepen themes. Similes, for example, directly compare two unlike things, often using "like" or "as" to create relatable (Kindree, 2022), concrete images that enrich a reader's comprehension and emotional engagement. Cognitive linguists argue that figurative language, including similes, shapes conceptual understanding, reinforcing how metaphorical thinking is integral to human cognition. However, analyzing similes across different languages, such as Arabic, presents unique challenges due to cultural and linguistic specificity (Ellis, 2024). Arabic similes often draw from distinct cultural references and rhetorical traditions, necessitating consideration of cultural context when interpreting them (Ryan, 2023). Ultimately, a multifaceted approach to analyzing figurative language—accounting for cognitive insights, cultural nuances, and the emotional impact on readers—enhances our appreciation of how these devices shape meaning across literary traditions.

Computational Approaches in Literary Studies

The evolution of simile detection in computational literary studies has advanced significantly, transitioning from rule-based approaches to sophisticated *machine learning* and deep learning methods. Initially, rule-based systems relied on identifying specific linguistic markers like “like” or “as,” which worked well for basic similes but struggled with complex figurative language (Fu, 2022). With increased computational power, statistical models emerged, allowing researchers to analyze large datasets and uncover patterns beyond immediate human perception. The advent of *machine learning*, particularly with neural networks like CNNs and BERT models, has further revolutionized the field. CNNs capture contextual information effectively, while BERT models excel in understanding nuanced language, making them highly suited for complex simile detection (Lai & Nissim, 2024). Hybrid models that combine rule-based and *machine learning* techniques, as well as multi-task learning frameworks, have further enhanced detection capabilities (Fu, 2022). However, challenges persist, such as the need for context sensitivity, cultural and linguistic adaptability, and handling ambiguity between literal and figurative language (Jimi & Shimada, 2024). Future research aims to refine models for a deeper comprehension of language and thought, integrating knowledge bases and advancing neural architectures. This progress in simile detection not only deepens literary analysis but also enriches our understanding of figurative language and cognition across diverse literary traditions.

Simile Detection in Arabic Texts: Challenges and Techniques

Simile detection in Arabic texts presents unique challenges and is a growing area of research interest due to the language’s complex morphological structure, regional dialectal variations, and the rich use of figurative language in Arabic poetry and literature. Arabic similes often involve intricate syntactic constructions, making automated detection difficult for current algorithms, which can struggle with accuracy and comprehensiveness. To address these challenges, recent studies have utilized Abdul-Raof’s linguistic framework, which decomposes similes into components such as the *mushabbah* (topic), *mushabbah bihi* (vehicle), *wajh al-shabah* (ground), and *adāt al-tashbih* (simile particle) (Humoud, 2020). This structural breakdown has improved the ability of algorithms to identify and categorize similes effectively. Furthermore, advances in natural language processing (NLP) techniques, including part-of-speech tagging, dependency parsing, and *machine learning* classifiers, have provided more nuanced tools for simile analysis (Elgabry et al., 2021). Notably, recent research by Al-Ghamdi et al. (2020) achieved a

high accuracy rate by using a hybrid rule-based and *machine learning* approach, while King Abdulaziz University’s annotated corpus of over 10,000 Arabic similes serves as a crucial resource for model training. Cognitive linguistic analyses, such as (Mahmoud & Zrigui, 2017) work on Quranic similes, offer additional insights into the conceptual structures underlying Arabic figurative language. While these innovations mark significant progress, capturing the full complexity and diversity of Arabic similes remains a challenge (Elgabry et al., 2021), with ongoing research promising further advancements through a combination of linguistic and NLP expertise (Abugharsa, 2022).

The existing literature on simile analysis has made substantial progress in identifying and interpreting figurative language, but notable gaps remain, particularly in the nuanced classification of similes within Arabic rhetoric. Current studies often lack a culturally informed framework to accurately capture the distinct structures and meanings of Arabic similes, which are deeply rooted in the language’s unique rhetorical and literary traditions. Furthermore, while computational methods have advanced in detecting similes, they frequently struggle with Arabic’s linguistic complexity, such as its morphology, syntactic structures, and regional variations. Additionally, many models focus on surface-level markers without accounting for the rich contextual cues and cultural resonances inherent in Arabic similes, leading to incomplete or inaccurate interpretations.

This study addresses these gaps by employing a comprehensive, interdisciplinary approach that integrates linguistic analysis, literary criticism, and computational techniques. By drawing on Arabic rhetorical frameworks and combining them with modern NLP tools, this research aims to offer a more precise classification of Arabic similes, accounting for both structural and contextual dimensions. This approach not only enhances computational accuracy but also bridges the cultural and cognitive aspects of simile interpretation, enriching our understanding of Arabic figurative language within a broader literary context. Ultimately, this study demonstrates the value of an interdisciplinary, culturally sensitive methodology, paving the way for future research that further integrates technology with linguistic and literary insights to capture the complexity of Arabic rhetorical expression.

Similes as Part of Arabic Rhetorical Features

Literature and everyday language frequently use similes, a figure of speech. A simile is a comparison between two things using the words “like” or “as.” However, many individuals struggle to differentiate between various

tropes such as metaphor, metonymy, simile, synecdoche, hyperbole, and polysemy. There is a lack of consensus among rhetoricians regarding the positioning of each trope within the broader arrangement of figures of speech (Literary Devices, 2020).

There has been ongoing debate about the nature of metaphors, but there is generally widespread agreement when it comes to similes. The American Heritage College Dictionary defines a simile as a figure of speech that compares two essentially different things using words like or as' (p. 1270). Similes, like other definitions found in dictionaries and rhetorical handbooks, highlight three key properties: (i) they involve a comparison; (ii) the comparison is made clear; and (iii) the comparison involves entities that are not typically considered comparable, making it metaphorical in some way (Ortony & Miller, 1993).

In Cognitive Grammar (Langacker, 1987), concepts are defined in relation to cognitive domains. The concept's domain matrix forms when each concept connects to a limitless range of domains. The concept of a "flower" encompasses not only its role as a physical component of a plant but also its involvement in contexts such as sexual reproduction, as a food source for insects, and as a decorative item in various settings Israel et al. (2004).

So, what factors influence the interpretation of similarities and differences? The interpretation of similes in cognitive grammar involves understanding the concept's domain matrix, where various domains can have varying levels of importance in relation to a concept. The prominence of different domains can vary depending on the specific situation in which the concept is being considered Israel et al. (2004).

Figure 1 visualizes this concept using VOSviewer (<https://www.vosviewer.com>). It maps out key terms related to similes in literature from the Web of Science database, covering papers from 2012 to 2023. This period encapsulates recent developments and ongoing research in the field of computational linguistics and simile detection. The results from this period reflect advancements in natural language processing, deep learning, and hybrid approaches, providing a representative sample of how simile recognition has evolved in recent years. The analysis focuses on titles containing keywords such as "SIMILE," "FIGURATIVE LANGUAGE," "LANGUAGE METAPHOR," and "FIGURATIVE MEANING." These keywords are grounded in the study's aim to explore a broad spectrum of figurative language devices that enhance rhetorical and emotional depth in texts. Similes and metaphors are central figures of speech in Arabic rhetoric, and their understanding is crucial to both traditional and modern linguistic analysis. Furthermore, the inclusion of "HUMOUR RECOGNITION" and "IRONY DETECTION" broadens the scope by considering how figurative language is used to convey humor and irony, which are often subtle

and complex forms of expression that challenge computational detection. By covering these dimensions, the study aims to comprehensively assess the capabilities of *machine learning* in detecting and interpreting various forms of figurative language across diverse contexts, from classical to contemporary literature (Al Zahrawi et al., 2024).

The visual map provided in Figure 1 offers a concentrated analysis of scholarly dialogue around "simile," illustrating its interconnectedness with a multitude of linguistic concepts. This analysis encapsulates research from 2012 to 2023, reflecting the enduring significance of similes in linguistic studies and the evolving inquiry into how this figure of speech shapes and is shaped by the intricate workings of human language and thought.

Categories of Similes

In terms of the two ends of simile, that is, the likened-to and the likened (Abdul-Raof, 2006, pp. 206–208) divided the types of similes into four categories:

- 1) *Perceptible-perceptible simile*: This is a simile in which the two ends correspond to one of the five senses, namely sight, hearing, smell, taste, and touch. As a result, both the likened-to and the likened are comparable since both are observable entities, as in: "Laylm's hair is like the night," where both the likened-to (the hair) and the likened (the night) are seen.
- 2) *Cognitive-cognitive analogy*: This category refers to similes with both ends devoted to cognition or emotional feelings: (i) cognition, rather than the five senses, is represented by expressions such as (- hope), (- luck), (- knowledge), (- ignorance), (- stupidity), (- intelligence), (- courage), (- good manners), (- politeness), and (- intelligence) (- an opinion). These terms, in other words, identify abstract nouns, as in: - Knowledge is like life.
(ii) Emotional states such as (- happiness), (- fear), (- wrath, rage), (- sadness), (- pain), and (- hunger), as in: - Hunger is torment. Both the likened-to (hunger) and the likened (suffering, punishment) are entities that communicate feelings.
- 3) *Cognitive-perceptible simile*: When one end of the simile is cognitive and the other is perceptible, as in: - Your ideas are like the darkness of the night. - Happiness is as sweet as honey. Whereas the likened nouns (- concepts, viewpoints) and (- contentedness) represent cognitive things, the likened nouns (- the darkness of night) and (- honey) reflect perceptible beings that can be seen, felt, and tasted.
- 4) *Observable-cognitive simile*: This is a type of simile in which the first end is visible while the other is cognitive, as in: - Your voice represents hope.

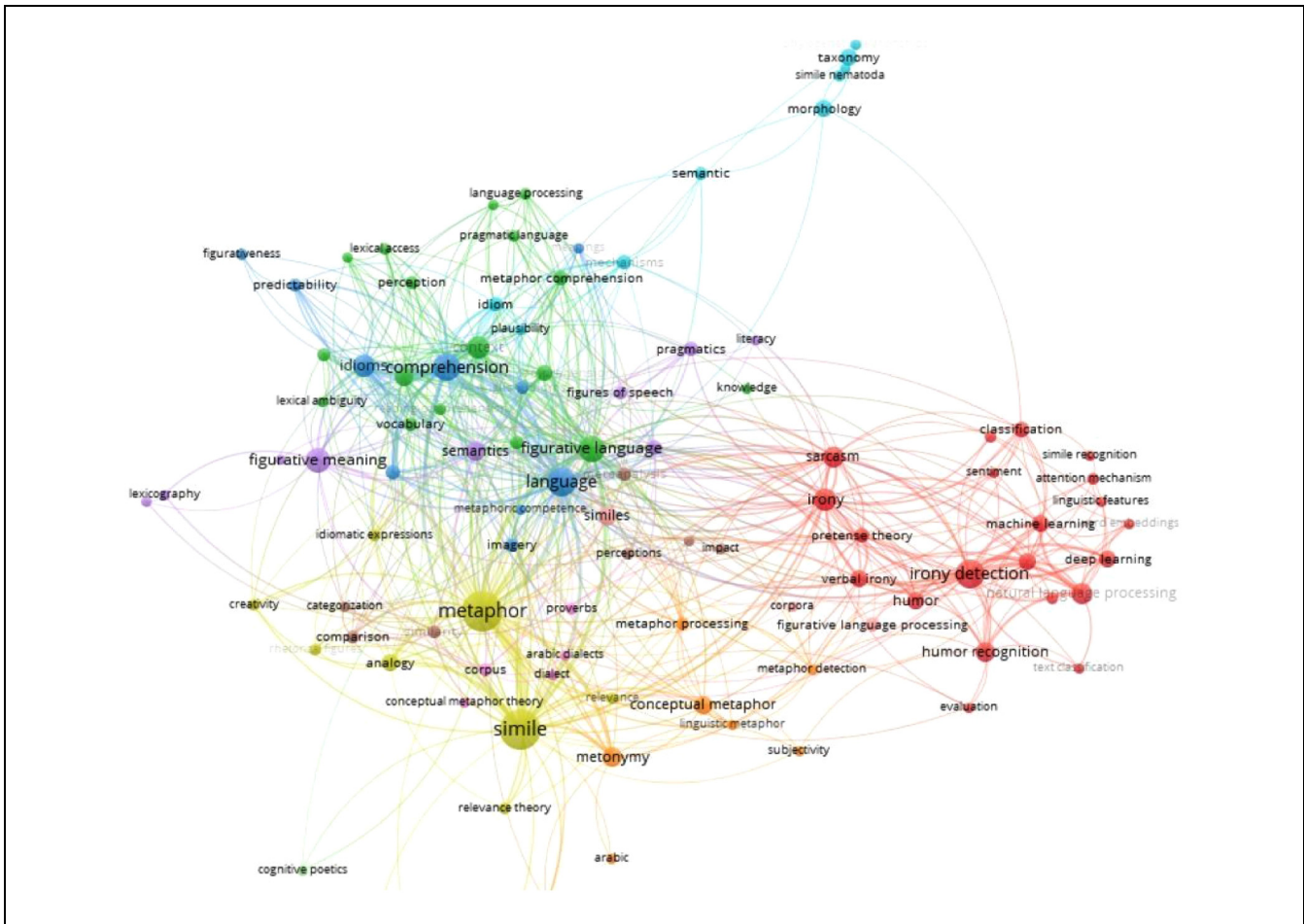


Figure 1. Keywords analysis on web of knowledge word extraction related to similes.

where the likened noun (- voice) represents a perceptible entity that can be heard and the likened-to noun (- hope) indicates a cognitive entity that can be experienced.

Simile Extraction

Simile extraction, a pivotal element in the broader field of linguistic analysis and natural language processing, involves the intricate task of identifying and deciphering similes in textual data. As stated above, similes, which are figurative comparisons typically constructed using “like” or “as,” serve to enrich the language, adding depth and nuance. However, their identification and analysis in computational terms come with challenges. The inherent diversity in simile construction and the heavily context-dependent nature of these figures of speech make automated extraction a complex endeavor. This task not only demands an understanding of linguistic structures but

also requires an appreciation of the subtleties and nuances embedded in human language.

Overview of Existing Techniques

This section investigates the array of methodologies and algorithms that have been developed and employed in the realm of simile extraction. The journey of simile extraction techniques is marked by an evolution from rule-based systems to more advanced *machine learning* and artificial intelligence approaches, each bringing unique strengths and facing distinct challenges.

Rule-Based Approaches. Early efforts in simile extraction primarily relied on rule-based systems. These systems were designed to identify similes by searching for explicit lexical patterns and markers, typically the words “like” or “as.” For example, a simple rule might look for structures matching the pattern “X is like Y” or “as X as Y.”

While effective in identifying clear, straightforward similes, these methods often struggled with more complex or subtly expressed similes. Additionally, they were limited by the need for extensive manual curation of rules and patterns, making them less adaptable to the vast and varied nature of natural language.

Statistical Methods. With the advent of statistical models in natural language processing, simile detection saw advancements in its capability to handle more varied and complex structures. These models, often leveraging large corpora of text, were able to discern patterns and associations that could indicate the presence of a simile. However, while more flexible than purely rule-based systems, statistical methods still faced challenges in accurately interpreting the nuanced and context-dependent nature of similes.

Machine Learning and Deep Learning Approaches. Recent developments in simile extraction have been driven by the application of machine learning, particularly deep learning techniques. These approaches, utilizing models such as neural networks, enable the learning of language representations directly from data, allowing for a more nuanced understanding of similes. Deep learning models can capture the subtleties of context and usage, making them more adept at identifying less obvious similes. Notably, techniques such as word embeddings and transformer-based models—like **BERT**, including the German BERT model developed by deepset (<https://www.deepset.ai/german-bert>), and **GPT**, accessible via ChatGPT (<https://chat.openai.com/>)—have shown promise in better capturing the complexities of figurative language. Despite their potential, these models require substantial computational resources and large annotated datasets for training, and they still face challenges related to linguistic diversity and idiomatic expressions.

Hybrid Approaches. Hybrid approaches that combine rule-based methods, statistical models, and *machine learning* have also emerged. These methods aim to leverage the strengths of each approach, using rules and patterns to guide the learning process and improve the precision of *machine learning* models. By integrating various techniques, hybrid systems seek to balance flexibility, accuracy, and computational efficiency.

Table 1 presents a comprehensive review of various methodologies and algorithms developed for simile extraction in linguistic analysis and natural language processing, highlighting specific description for each reference.

The exploration of figurative language, particularly similes, in text has seen a variety of approaches, each with its own merits and limitations. Rule-based methods, such as Fass's Met* Algorithm and Zeng et al.'s work on idioms, rely on predefined pattern libraries to identify figurative language, offering a more straightforward and interpretable framework. However, their effectiveness is closely tied to the comprehensiveness of the pattern library. Statistical methods, exemplified by Veale and Hao's and Li et al.'s research, harness large-scale data to identify patterns, offering more adaptability but requiring extensive corpus data for effective training. *Machine learning* and deep learning approaches, like those by Rei et al. and Devlin et al., leverage neural networks and models like BERT for nuanced language understanding, but they demand substantial computational resources and large datasets.

Hybrid approaches, including Shutova's, Mason's, and Ren and Zhang's models, attempt to combine the strengths of rule-based and statistical methods, aiming for both interpretability and adaptability. However, these too can be complex and resource-intensive.

Given these considerations, and recognizing the limitations of our corpus data, a rule-based method appears to be the most pragmatic choice for our current needs. While this approach may not capture the full nuance of language that *machine learning* methods offer, it provides a reliable, less resource-intensive way to identify similes, making it a sensible starting point for our analysis. This choice allows for a more controlled and manageable exploration of figurative language, which is particularly beneficial when dealing with limited data or seeking specific, pattern-based insights.

Research Methodology

This section presents the findings of the simile extraction and classification process applied to the *Al-Abrat* corpus using a Python-based algorithm. The results are divided into two main categories: confirmed similes and compound similes. We will also discuss the manual validation process used to ensure accuracy and the relationship between the computational approach and literary analysis. Finally, this section will conclude by revisiting the research hypotheses and addressing the limitations of the study.

Theoretical Framework

The theoretical framework for this study is grounded in Abdul-Raof's (2006) categorization of similes, which provides a detailed taxonomy for analyzing rhetorical comparisons, including perceptible-perceptible, cognitive-cognitive, and mixed types. This framework is

Table 1. Various Methodologies and Algorithms Developed for Simile Extraction in Linguistic Analysis and NLP.

Approach type	Specific approach	Primary researcher (s)	Description	References
Rule-based approaches	Fass's Met* Algorithm	Dan Fass	Identifies figurative language by matching text to a predefined pattern library.	Fass (1991)
	Zeng et al.'s Work on Idioms	Zeng et al.	Developed a rule-based system for idiom recognition, which can be adapted for simile detection.	Ziheng and Bhat (2021)
Statistical methods	Veale and Hao's Simile Recognition	Tony Veale and Yanfen Hao	Implemented a statistical approach to identify creative similes in text, using large-scale web mining.	Veale and Hao (2007)
	Li et al.'s Corpus-Based Study	Li et al.	Utilized statistical methods to identify similes in a Chinese corpus, focusing on frequency of simile markers.	Liu et al. (2022)
Machine learning and deep learning approaches	Rei et al.'s Neural Network Model	Rei et al.	Employed a neural network approach for metaphor detection, adaptable for similes, focusing on context and word embeddings.	Rei et al. (2020)
	Devlin, J et al.'s BERT-Based Model	Devlin et al.	Used a BERT-based model for understanding figurative language, effective in detecting nuanced similes.	Devlin et al. (2018)
Hybrid approaches	Shutova's Hybrid Model for Metaphor Identification	Ekaterina Shutova	Combined rule-based and statistical methods for metaphor identification, applicable to similes.	Shutova (2015)
	Mason's Simile and Metaphor Analysis Tool	Zachary Mason	Integrated rule-based patterns with <i>machine learning</i> for analyzing similes and metaphors in literary texts.	Mason (2004)
	A Hybrid Representation-Based Simile Component Extraction	Ren, D., Zhang	The paper presents the "Hybrid Representation-based Component Extraction" (HRCE) model for effectively extracting simile components in sentences by utilizing multi-level word, concept, and character representations, significantly improving upon existing methods in natural language processing.	Ren et al. (2020)

operationalized using computational linguistics, specifically Python-based algorithms and regular expressions, to automate the detection and classification of similes in *Al-Abrat*. A hybrid approach integrates manual validation to address nuanced or unconventional similes, ensuring accuracy and depth. Together, these elements provide a systematic, replicable, and nuanced methodology for analyzing Arabic rhetorical devices in literary texts. Figure 2 summarizes the theoretical framework.

Data of the Study

Although this study identified 200 similes within *Al-Abrat*, only 9 were chosen for detailed analysis to ensure a thorough, focused examination that aligns with the study's objectives. This selection provides several key benefits:

- *Representative variety*: The nine chosen samples represent the diversity of simile types found across

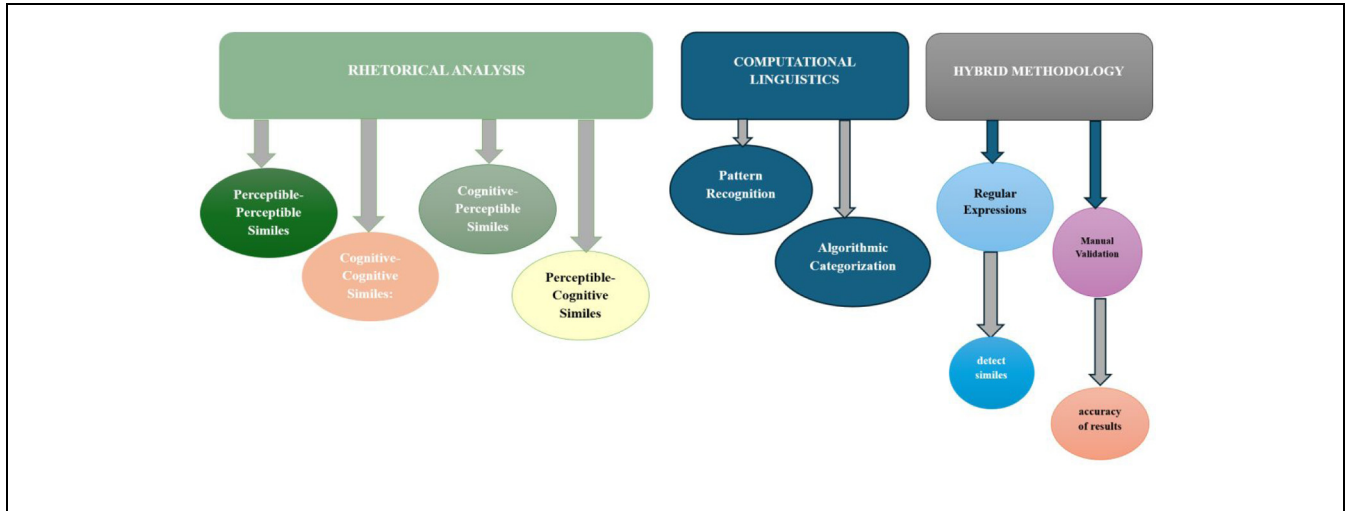


Figure 2. Theoretical framework.

the full dataset. By focusing on this subset, the study can highlight different rhetorical structures and cognitive impacts that exemplify the broader patterns within the text without redundancy.

- *Depth of analysis:* A detailed examination of each sample allows the researchers to explore each simile's linguistic, rhetorical, and cognitive dimensions comprehensively. This level of analysis would be impractical across all 200 samples, as it would risk diluting the clarity and depth of insights achieved in the discussion.
- *Clarity and accessibility:* Selecting a smaller number of samples balances comprehensiveness with readability, making complex findings more accessible to readers. This approach allows for nuanced discussion without overwhelming readers with extensive data points.

Thus, focusing on these nine representative samples enables the study to present well-supported conclusions about the structure and function of Arabic similes in *Al-Abrat* while maintaining analytical rigor and clarity.

Simile Detection Results

The methodology in this research follows a two-step process: first, similes are detected in the text using a Python-based algorithm, and second, these similes are categorized according to Abdul-Raof's four types, ensuring both computational and theoretical alignment. Initially, similes are extracted from the specified text utilizing a Python-based algorithm program. Subsequently, these similes are categorized into distinct types, aligning with the four categories delineated by Abdul-Raof (2006). This approach ensures a structured analysis, facilitating

the evaluation of the program's effectiveness in simile identification and classification.

Simile Classification

The methodology is based on a quantitative and qualitative approach using an algorithm to extract similes from the translated English version of *Al-Abrat*. In the exploration of simile classification, Abdul-Raof (2006, pp. 207–208) delineates 14 distinct types of similes, such as:

- *Single simile:* Straightforward, as exemplified by the comparison of advice to medicine.
- *Multiple similes:* More complex, where the system compared the rhetorical portrayal of a president and his predecessor by emphasizing their respective negative traits.
- *Compound simile:* Combined elements, likening the sun to a golden ball.
- *Synopsis simile:* Reduces the comparison to a simple equivalence, such as a speech to honey.
- *Detailed simile:* Goes further, specifying the aspect of comparison, like the sweetness of honey to a speech.
- *Unrestricted simile:* No limit to the scope of comparison, for example, advice to gold.
- *Confirmed simile:* Strengthens the comparison, for example, equating advice directly to gold.
- *Perceptible-perceptible simile:* Relies on tangible comparisons, for example, juice to seawater.
- *Cognitive-cognitive simile:* Deals with abstract concepts, for example, hunger to despair.
- *Cognitive-perceptible and perceptible-cognitive simile:* Blend abstract and tangible elements, for

example, luck to flour on a windy day and perfume to happiness.

Simile Detection Algorithm

Drawing on the comprehensive classification by Abdul-Raof (2006, pp. 207–208) the research team devised a structured approach for evaluating the capability of a Python-based algorithm to identify similes within a literary corpus. This corpus comprised the English translation of “Alabrat.” The designed structure is rooted in the nuanced categories outlined by Abdul-Raof, providing a robust framework for computational analysis.

By systematically aligning the program’s detection algorithms with the diverse types of similes—ranging from simple to complex, tangible to abstract—the researchers aim to rigorously test the use of a programming language such as Python form’s proficiency in recognizing and interpreting these rhetorical devices within the translated text. The method used is based on using regular expressions that would match the simile category.

Regular Expression Patterns for Simile Detection. The following regular expression patterns were used for the compound similes, confirmed similes and single simile, and cognitive-perceptible similes.

```
patterns = {
    "compound_similes": r'\b\w+\b\s+(is|am|are|was|were)\s+(as|like)\s+\b\w+\b\s+\b\w+\b',
    "single_similes": r'\b\w+\b\s+(is|was|am|are|were)\s+(as|like)\s+(?:\b\w+\b\s+)*\b\w+\b',
    "confirmed_similes": r'\b\w+\b\s+(is|was|am|were)\s+\b\w+\b',
    "cognitive_perceptible_similes": r'\b\w+\b\s+(is|was|am|were)\s+(as|like)\s+\w+\s+in\s+\w+\s+\b\w+\b',
}
```

Data Processing and Output. After establishing the categories of similes and designing the regular expressions, the next step involved processing the text to identify and extract similes. The Python-based algorithm was tasked with analyzing the text to detect specific patterns associated with similes, allowing for an efficient extraction process.

The regular expressions were applied to pinpoint patterns indicative of similes within the text. The designated pattern for identifying similes was: “noun/phrase + is/am/are/was/were + as/like + noun/phrase.” The Python-based algorithm processed the text sentence by sentence. Initially, the text was extracted from a PDF file and stored in a database as individual sentences. The

program then searches for matches based on a set of regular expressions defined in the *patterns* dictionary (set of regular expressions tagged by simile category). When a match is found, the sentence containing the simile is automatically saved to an Excel file (duplicates are ignored).

The Excel file was structured to categorize sentences based on various types of similes, as defined by the regular expressions listed in the Python *patterns* dictionary. Additionally, instances containing terms such as “dislike” and “don’t like” were systematically excluded from retrieval. The identified similes were compiled into a list, enabling subsequent analytical procedures.

Exclusion of Non-Similes. Moreover, instances containing terms such as “dislike” and “don’t like” were systematically excluded from retrieval. Similes identified were compiled into a list, enabling subsequent analytical procedures.

Classification of Similes. The extracted similes were systematically classified into specific types, adhering to the framework of the four simile categories as established by Abdul-Raof (2006). The developed Python-based algorithm was employed for this task, designed to sort the similes into four nuanced categories: perceptible-perceptible simile, cognitive-cognitive simile, cognitive-perceptible simile, and perceptible-cognitive simile. This classification leveraged the descriptive power of language, bridging the tangible with the abstract and enabling a refined analysis of simile use within the text.

Algorithmic Approach to Simile Classification

The second step of the research falls into the process of classifying similes. An algorithmic approach was adopted where each simile was meticulously evaluated through a series of conditional if-else statements aimed at determining the nature of the elements being compared. This algorithmic classification was pivotal in the nuanced analysis of the simile structure.

- *Perceptible-perceptible similes:* Both elements are tangible and can be experienced through the five senses, drawing a parallel between two sensory experiences (e.g., sight or touch).
- *Cognitive-cognitive similes:* Both elements are intangible, related to mental processes or emotional states (e.g., concepts, thoughts, or feelings).
- *Cognitive-perceptible similes:* These similes juxtapose a sensory experience with a cognitive or emotional one.

- *Perceptible-cognitive similes*: One element is a cognitive entity and the other a visible, tangible experience.

Back Translation

Back-translation (BT) is a valuable quality control method in translation that involves translating text from the target language back into its original source language, enhancing accuracy, especially in fields requiring precision. Rather than producing an alternative translation, the purpose of BT is to act as a “mirror,” reflecting the content of the translation to identify any discrepancies or inconsistencies (Mirela, 2024). The BT process includes several steps: translating the original text into the target language, having a different translator retranslate the target text into the source language, comparing the original with the back-translation, analyzing discrepancies, and revising as necessary (Górniak-Woźniak, 2023). BT proves particularly useful in highly regulated industries—such as medical, legal, financial, and technical fields—where it adds a layer of quality control, reduces the risk of translation errors, and addresses culturally sensitive aspects in marketing and branding (Mirela, 2024). Additionally, BT serves as a tool to assess machine translation accuracy and to reveal cultural nuances (Edunov et al., 2020). However, BT can be resource-intensive and depends heavily on translator skill, as some errors in the forward translation may go unnoticed (Whatley, 2024). To maximize BT’s effectiveness, best practices include engaging native speakers, providing context, using multiple translators, considering cultural differences, verifying terminology consistency, and avoiding exclusive reliance on machine translation. Despite its limitations, BT remains a valuable quality assurance method, particularly for complex linguistic elements like metaphors, idioms, and slogans (Samson, 2019).

Results

This section presents the findings of the simile extraction and classification process applied to the *Al-Abrat* corpus using a Python-based algorithm. The results are divided into two main categories: confirmed similes and compound similes.

Simile Detection Results

In the comprehensive analysis of *Al-Abrat*, our research leveraged a Python-based algorithm to meticulously parse the text for literary similes in its translated version into English. This advanced detection method yielded a substantial finding: 212 similes within the text.

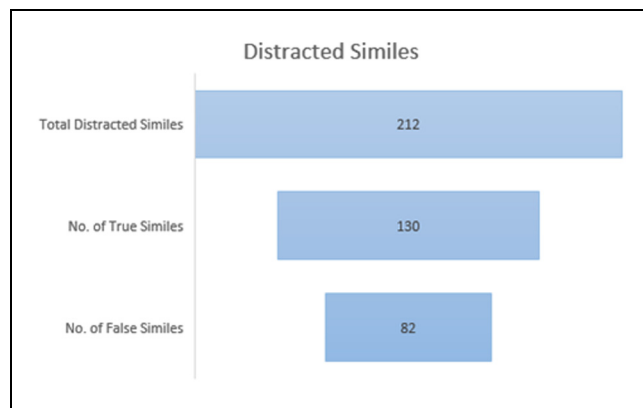


Figure 3. Manual validation of detected similes.

Categorization of Similes

Our analysis discerned two distinct categories of similes: confirmed similes and compound similes.

- *Confirmed similes*: Characterized by their straightforward comparative nature, confirmed similes directly align one element with another, exemplifying clarity and simplicity. For instance, *Al-Abrat* offers vivid examples, such as likening fairness to the pristine drop on a sunlit lily, illustrating this technique with eloquence.
- *Compound similes*: In contrast, compound similes emerged as a more intricate construct, engaging multiple elements to draw parallels, thereby offering a layered understanding of the subjects compared. For example, similes that juxtapose the human heart with both a “stationary reservoir” and a “powerful sword” paint a dualistic picture of human nature’s tranquility and ferocity.

Manual Validation of Simile Detection

A subsequent manual review of the algorithm’s output was crucial in validating the accuracy of the simile detection. After going through this thorough process, 130 of the similes were found to be true to the literary definition, while 82 were found to be false positives (Figure 3). This emphasizes the necessity of human oversight in computational literary analysis.

Examples of True Simile and Non-simile

The following are examples of true similes detected in the text (Tables 2 and 3):

This simile (Sample 2) compares the speaker’s voice to the rumble of thunder, using “like” to create an image of a powerful, commanding tone. The simile emphasizes

Table 2. Sample 1—True Simile (Noun | Pronoun + Verb + Like + Noun).

English version	Arabic original simile	Simile structure	Back translation
He stood like a statue.	فجى أرملة في مكانه تجود الضم	Noun + is like + Noun	وقف مثل تمثال

Table 3. Sample 2—True Simile (Noun + Like + Noun Phrase).

English version	Arabic original simile	Simile structure	Back translation
It started speaking with a voice like a thunderbolt of lightning in the horizon.	ثم أخذ ينطق بصوت كأنه جلبة الرعد في آفاق السماء	Verb to be + an adjective	He began to speak with a voice that sounded like the rumble of thunder in the sky

Table 4. Sample 3—True Simile (As + If + Pronoun + Verb To Be).

English version	Arabic original simile	Simile structure	Back translation
She was staggering in her walk as if she was walking upon uneven sand.	تترع في مشيتها كأنها تمشي على رملة ميثاء	As + if + pronoun + verb to be	She staggers in her walk as if she is walking on Metha sand.

Table 5. Sample 4—True Simile (As + Noun).

English version	Arabic original simile	Simile structure	Back translation
like the heart of one who enters a racecourse as a competitor, spending on it all his wealth with no knowledge whether he shortly will win or lose, become the most fortunate or the most wretched.	ما يكون في نفس اللهايب إلى ميثاء سبابة قد خاطر فيه بجميع ما يملك فهو لا يعلم أكون بعد ساعة أسعد لناس أم أشقاهم ؟	As + noun	What happens to someone who goes to a racing track and has risked everything he owns, not knowing whether after an hour he will be the happiest or most miserable of people?

the forcefulness of the voice, capturing its impact on the listeners as something both felt and heard, much like a thunderbolt electrifies the sky (Table 4).

This simile (Sample 3) compares her unsteady movement to someone walking on uneven sand, emphasizing the difficulty and instability of her steps, as if the ground beneath her was shifting and unreliable (Table 5).

The simile in Sample 4 compares the emotional state of a person entering a risky venture to that of a competitor at a racecourse, conveying the suspense and high stakes involved (Table 6).

The sentence “A woman like this should not be permitted to live” uses the word “like” to refer to a type of person or behavior but does not draw an illustrative comparison between two different things, making it a non-simile (Table 7).

The sentence “A person like me cannot endure this kind of behavior” uses the word “like” to compare

tolerance or character rather than creating an imaginative or illustrative comparison. Therefore, it does not serve as a simile (Table 8).

The phrase uses a contrastive structure rather than a simile. “Likewise” signals a parallel, highlighting the opposing states (fortunate and unfortunate) rather than comparing two entities figuratively. This rhetorical contrast emphasizes the dual nature of human experience under the influence of life (Table 9).

The sentence “Women like this deceive men like you; filthy women like this do not have the hearts to love with” does not form a simile. It uses the word “like” to categorize individuals based on perceived behavior, rather than drawing a figurative (Table 10).

The sentence “It continued to mount down until there remained between it and the earth what appeared like a mile or even less in distance” uses “like” to describe an approximation of distance.

Table 6. Sample 5—Non-Similes (Noun + Like + This + Verb [Auxiliary]).

English version	Arabic original simile	Simile structure	Back translation
A woman like this should not be permitted to live.	مَا كَادَ لَهَا أَنْ يَحْيَا عَلَى مِثْلِ ذَلِكَ	Noun + like + this + verb (auxiliary)	Someone like her could not live like that.

Table 7. Sample 6—Non-Similes (Noun + Like + Personal Pronoun + Verb [Auxiliary]).

English version	Arabic original simile	Simile structure	Back translation
A person like me cannot endure this kind of behavior, but I did, until he came to me at the beginning of one night in a stressful state with a severe problem.	مُخَلَّلاً فِي سَبِيلِ ذَلِكَ مِنْ حُمَةِ وَشَوَاسِهِ وَفُسَادِ تَصَوُّرَاتِهِ وَغَرَابَةِ أَطْوَارِهِ مَا لَا طَلْقَةَ لِيَلِي بِاخْتِلَالِ بَيْتِهِ حَتَّى جَاءَنِي نَدَا لَيْلَةٍ بِدَاهِيَةِ الدَّوَاهِيِ وَنُصْبَةِ الْمُصَافِ فَكَانَتْ . آخِرَ عَهْدِهِ بِهِ .	Noun + like + personal pronoun + verb (auxiliary)	Possibly for the sake of This is his foolishness, obsessiveness, corrupt notions, and eccentricities, which someone like me cannot bear. With such endurance, until he came to me one night with the onset of treachery and the affliction of misfortunes, and it was my last covenant with him.

Table 8. Sample 7—Non-Similes (An Adverb “Likewise”).

English version	Arabic original simile	Simile structure	Back translation
How man, through the hands of life, can become fortunate and likewise unfortunate?	!!فَمَا أَسْعَدَ الدَّهْرَ بِالْإِنْسَانِ وَمَا أَشَقَى الْإِنْسَانَ بِهِ	Likewise is mistaken for “like.”	How happy life is for man and how miserable man is during it!!

Table 9. Sample 8—Non-Similes (Noun + Like + Personal Pronoun + Verb [Auxiliary]).

English version	Arabic original simile	Simile structure	Back translation
If I were to leave her, I would be committing such a big sin that its regret would not part from my conscience till death. His father replied, “Women like this deceive men like you, filthy women like this do not have the hearts to love with.”	وَأَحْسِبُ أَنِي إِذَا فَارَقْتُهَا قَتَلْتُهَا، وَجَنَيْتُ عَلَيْهَا جُنَايَةَ لَا يَفَارِقُنِي النَّدَمُ عَلَيْهَا حَتَّى الْمَوْتِ قَالَ: ذَلِكَ مَا يَجْدَعُ بِهِ أَمْثَالُهَا أَمْثَالُكَ، فَلَيْسَ لِلنِّسَاءِ الْعَاهِرَاتِ قُلُوبٌ يَجِبْنَ بِهَا	Pronoun + would + be + would be (auxiliary)	If I were to leave her, I would be committing such a big sin that its regret would not part from my conscience till death. His father replied, “Women like this deceive men like you; filthy women like this do not have the hearts to love with.”

Discussion of Results

This section provides a detailed interpretation and analysis of the similes extracted from the *Al-Abrat* corpus, classified using a Python-based detection. By analyzing true similes within classical Arabic literature, we validate our hypotheses regarding their structural and functional

roles. Specifically, this analysis provides insights into how similes convey emotional and cultural meanings unique to Arabic rhetoric, thereby enhancing our understanding of their linguistic complexity and cognitive impact on readers. The discussion will focus on

Table 10. Sample 9—Non-Similes (Noun+Like+Personal Pronoun+Verb [Auxiliary]).

English version	Arabic original simile	Simile structure	Back translation
It continued to mount down until there remained between it and the earth what appeared like a mile or even less in distance.	ازداد هبوطه ، حتى إذا لم يبق بينه وبين الأرض إلا ميل أو بعض ميل	Pronoun + would + be + would be (auxiliary)	His descent increased until only a mile or more remained between him and the ground.

illustrating how the identified similes contribute to the emotional and rhetorical depth of the text while also addressing the limitations of the detection algorithm, which encountered challenges in distinguishing between figurative and non-figurative uses of “like” and “as.” The section will conclude by revisiting the research hypotheses and addressing the limitations of the study.

Discussion of Similes (Sample 1–9)

The similes and non-similes identified in this study provide valuable insight into the various ways figurative language is employed in Al-Abrat. Through careful analysis, we can discern the distinct rhetorical effects these figures of speech create, as well as the limitations of the detection algorithm in recognizing figurative intent.

Sample 1. For Sample 1, the sentence “He stood like a statue” is a simile because it compares two different things—a person standing and a statue—using the word “like” to draw the comparison. The purpose of this simile is to highlight the stillness or immobility of the person by likening it to the permanent stillness of a statue. A statue is an inanimate object, typically not capable of movement, that often represents a frozen moment or a figure poised in a permanent state. When we say someone “stood like a statue,” we imply that the person is so still that they could be mistaken for a statue, or that they are intentionally holding a very still pose, unmoving, possibly suggesting they are deep in thought, shocked, or deliberately remaining motionless for some reason.

Sample 2. For Sample 2, the effectiveness of the simile lies in its ability to create a vivid picture of the person’s stance by comparing it to something universally understood to be motionless. This allows the reader or listener to immediately grasp the degree of stillness of the person standing. As the silence of the room was abruptly broken, it began to speak in a voice that resonated with the commanding presence of nature’s own fury. The timbre of its speech struck the air like a thunderbolt of lightning in the horizon, a simile that perfectly encapsulated the sudden and arresting force with which the words were

delivered. This voice did not just reach the ears of those present; it seemed to ripple through them, carrying the weight and awe of a storm’s first flash across a darkening sky. The comparison, drawn with the use of “like,” vividly painted the speaker’s tone as something not just heard but felt, reminiscent of how a lightning strike can momentarily illuminate an entire landscape and leave a lasting impression on its witnesses. The power behind the voice was indeed as palpable and as electrifying as the natural phenomenon it was likened to, leaving no doubt about the importance of the message it carried.

Sample 3. For Sample 3, the simile compares the subject’s unsteady movement to someone walking on uneven sand, highlighting the difficulty and instability. Her movement was an unsteady ballet, a poignant display of struggle as she navigated the space before her. With each tentative step, she seemed to contend with an invisible, shifting terrain, her body swaying unpredictably as she advanced. The simile that came to mind was that she was staggering in her walk as if she were walking upon uneven sand. This comparison conveyed the essence of her challenge—the way her feet seemed to sink slightly, seeking stability where there was none, her balance as precarious as if she traversed the constantly changing surface of a windswept beach. The room’s solid floor beneath her offered no clues to her halting gait; it was as though her legs bore the memory of a dune’s inconsistency, adapting to a phantom unevenness that her eyes told her was not there. Each step was a testament to her determination, a delicate dance of will against the unseen trials her body faced.

Sample 4. In Sample 4, the simile draws a vivid comparison between the emotional state of a person stepping into a venture and a competitor at a racecourse. As he stepped into the arena of his latest venture, his heart was alight with the same tumultuous blend of hope and trepidation that might grip the soul of a novice gambler stepping onto the hallowed grounds of a racecourse. Like such a competitor, he had invested not just his funds but his very dreams into this uncertain endeavor, laying the entirety of his wealth upon the capricious altar of fate.

His every heartbeat was a drumroll that matched the thunder of hooves on the track, embodying the suspense that hangs in the air as the race begins—a suspense thick with the possibility of triumph or the dread of despair. With no foresight to comfort him, he was acutely aware that the outcome could swing with equal ease towards boundless fortune or ruinous loss. Each moment that passed was a stride made by the steeds of destiny, racing towards a finish line that would declare him either the most fortunate among men, basking in the glow of success, or the most wretched, enveloped by the shadow of loss. The simile underscores the extreme stakes and the intense emotional experience of someone fully committed to a venture, with all to gain or all to lose, suspended in the breathless in-between of not knowing.

Sample 5. In Sample 5, the sentence “A woman like this should not be permitted to live” does use the word “like,” but it does not serve the purpose of a simile, which is to compare and show similarity between two different things in a descriptive way. Instead, “like” in this context is used to refer to a type of person or a kind of behavior that the speaker thinks is objectionable. There is no illustrative comparison being made here; the word “like” functions to specify a category or type rather than draw a comparison for descriptive effect. In other words, for a simile to be present, the sentence would need to draw a comparison that enlightens or illustrates a characteristic by likening it to something else, such as “She is like a rose, beautiful but thorny,” which compares a woman to a rose to comment on her beauty and potential to cause pain. The distinction can be subtle, and it’s not solely the presence of words like “like” or “as” that defines a simile; it is how they are used to craft a comparative image or idea that matters.

Sample 6. In Sample 6, in the sentence “A person like me cannot endure this kind of behavior,” the word “like” is not employed to create a simile but rather to indicate a comparison of similarity in tolerance or character. The phrase categorizes the speaker within a group of individuals with a presumed similar endurance level, lacking the illustrative, descriptive purpose of a simile, which typically draws an imaginative connection between two disparate entities. Hence, despite the presence of “like,” the sentence does not serve the figurative function of a simile, as it does not aim to evoke a vivid picture or a deeper understanding through comparison.

Sample 7. In Sample 7, the phrase “How man, through the hands of life, can become fortunate and likewise unfortunate?” reflects on the dual nature of human existence and the role of fate or circumstances. It

contemplates the notion that life, personified as having hands, can bestow fortune or misfortune upon individuals. This metaphorical expression encapsulates the unpredictable journey where a person might experience success, happiness, and prosperity at certain times, which can be seen as the touch of life’s “fortunate” hand. Conversely, the same individual may encounter challenges, sorrow, or adversity, as if life’s “unfortunate” hand has been dealt. This duality underscores the randomness and caprice of existence, suggesting that the same forces that contribute to one’s rise can also contribute to one’s fall. It’s a philosophical inquiry into the nature of destiny and the human condition, acknowledging that while people may have agency, there are elements beyond control that can swing the pendulum of fortune in either direction.

Sample 8. In Sample 8, the sentence “Women like this deceive men like you; filthy women like this do not have the hearts to love with” is not a simile because it does not use “like” to directly compare two distinct things for the purpose of illustrating a shared quality or characteristic. Instead, in this context, “like” is being used to categorize or typecast certain women and men based on perceived behavior or attributes. In a simile, “like” serves to draw a comparison that typically connects an abstract concept with a concrete image to create a vivid description (e.g., “Her smile was like the first day of spring”). In the provided sentence, however, “like” is not setting up such a figurative comparison. It is simply being used in a literal sense to group women and men who are viewed as having similar characteristics or behaviors (deceptive women and the men they deceive). Furthermore, the use of “like” in the phrase “filthy women like this” functions as a descriptor to define a subset of women the speaker considers to be deceitful—it does not liken them to something else to illuminate a trait in a metaphorical way. The sentence is making a statement about reality as perceived by the speaker rather than employing a literary device for illustrative purposes. The sentence is making a statement about reality as perceived by the speaker rather than employing a literary device for illustrative purposes.

Sample 9. In Sample 9, the sentence “what appeared like a mile or even less in distance” conveys an approximation rather than a simile. It describes an object descending toward the earth, getting closer until there is only a small distance separating them. The phrase gives the impression of an observer estimating the gap between the object and the earth’s surface. The word “like” here is not used to draw a figurative comparison but to indicate uncertainty in the measurement. The sentence describes a literal scenario rather than employing a simile, as it

focuses on approximating the physical distance rather than highlighting any shared qualities between two different entities.

Findings

The results obtained from the Python-based simile detection algorithm demonstrate its overall effectiveness in identifying and classifying similes within the text of *Al-Abrat*. The extraction of 212 similes, followed by manual validation of 130 true positives, underscores the program's capability to handle a diverse range of simile structures, from simple, straightforward comparisons (e.g., "He stood like a statue") to more complex figurative language (e.g., "The heart of a competitor racing to win or lose"). This validates *Hypothesis 1*, which posited that the algorithm would successfully detect and classify similes with high accuracy.

The classification based on Abdul-Raof's framework also confirmed the diverse nature of similes used in *Al-Abrat*, ranging from perceptible-perceptible similes (e.g., comparisons of physical objects) to cognitive-cognitive similes (e.g., abstract comparisons like hunger to despair). These findings align with *Hypothesis 2*, which suggested that simile usage in the text would follow distinct patterns reflective of classical Arabic literature's cognitive and rhetorical functions.

The results obtained from the Python-based simile detection algorithm demonstrate its overall effectiveness in identifying and classifying similes within the text of *Al-Abrat*. The extraction of 212 similes, followed by manual validation of 130 true positives, underscores the program's capability to handle a diverse range of simile structures, from simple, straightforward comparisons (e.g., "He stood like a statue") to more complex figurative language (e.g., "The heart of a competitor racing to win or lose"). This validates *Hypothesis 1*, which posited that the algorithm would successfully detect and classify similes with high accuracy.

The classification based on Abdul-Raof's framework also confirmed the diverse nature of similes used in *Al-Abrat*, ranging from perceptible-perceptible similes (e.g., comparisons of physical objects) to cognitive-cognitive similes (e.g., abstract comparisons like hunger to despair). These findings align with *Hypothesis 2*, which suggested that simile usage in the text would follow distinct patterns reflective of classical Arabic literature's cognitive and rhetorical functions.

Overall, these findings indicate that while the rule-based approach used in this study is effective, it is not infallible. The combination of computational methods with manual validation proved essential for achieving accuracy and interpretative depth. This reflects the need

for further refinement of the algorithm or the possible integration of more advanced *machine learning* models that could handle subtleties and nuances in figurative language more efficiently. The study addresses a critical gap in Arabic rhetoric research by developing an NLP-based framework specifically for simile classification and cognitive interpretation, offering a novel approach that has not been explored in previous studies on Arabic literary devices. Previous works, highlighted by scholars like (Al-Mubarak & Hill, 2023) and (Al-Sharafi, 2004), have broadly addressed figurative language but often overlooked the distinct subtleties and complexities inherent in similes. Building on the foundational work of (Abdul-Raof, 2006) in categorizing similes, the research innovatively combines cognitive linguistic theories with computational analysis (Feldman, 2004). This novel, multidisciplinary approach stands out in the field, venturing into areas that have remained largely unexplored in the existing body of literature. This fusion of disciplines not only provides a more nuanced understanding of similes in Arabic texts but also opens up new pathways for future research in the realms of both linguistics and computational studies.

The strengths of the study are multifaceted and significant. First, its comprehensive analysis, merging linguistic scrutiny with computational techniques (Heaton et al., 2023), provides an in-depth understanding of simile structures in Arabic texts, a notable advancement in the field. This thorough approach allows for a more detailed examination than either discipline could achieve alone. Second, the practical application of the research is evident in the development of algorithms for simile extraction. These algorithms hold great potential for application in natural language processing, possibly enhancing the capability of machines to understand and process figurative language. Lastly, the research serves as a bridge between literary criticism, cognitive linguistics, and computational linguistics. By integrating these disciplines, the study contributes to a more rounded and thorough comprehension of Arabic rhetoric.

However, there are weaknesses to consider. The study's reliance on rule-based methods for simile extraction, while practical and less resource-intensive, may not capture the nuanced understanding that advanced *machine learning* methods offer. This could limit the ability to adapt to the varied and complex nature of similes in Arabic texts. Additionally, the manual review process, a crucial step in validating the output of the algorithm, introduces an element of subjectivity (Gurus, 2021). This subjectivity could potentially affect the accuracy and objectivity of simile classification, as the personal interpretations and biases of the reviewers might influence the final analysis.

The study contributes substantially by refining simile detection methodologies in NLP, adding depth to literary interpretations in Arabic criticism, and advancing cognitive linguistic theories on metaphor and simile processing. By providing a more detailed understanding of similes, the research enhances the appreciation of stylistic choices and thematic depth in these texts. This new lens allows for a richer, more nuanced engagement with Arabic literature, potentially leading to more profound interpretations and discussions within the literary community.

In the realm of cognitive linguistics, the study significantly contributes to the understanding of how the human mind processes figurative language, with a particular focus on similes. This insight is vital for comprehending the cognitive mechanisms underpinning language comprehension and emotional responses to literature. It broadens the scope of research in cognitive linguistics, offering valuable data on the interaction between language and thought.

Moreover, in the field of natural language processing, the research provides crucial data and methodologies for the development of algorithms capable of detecting and interpreting similes in Arabic texts. This contribution is particularly valuable for advancing computational linguistics, as it addresses the complex challenge of processing figurative language—a task that machines have traditionally struggled with. By improving these capabilities, the study aids in the creation of more sophisticated language processing tools, enhancing the interaction between humans and technology in language understanding.

The study of Arabic rhetoric's similes has limitations, including a singular corpus like *Al-Abrat* and potential algorithmic bias. It suggests incorporating diverse texts for a more representative understanding. The manual review process introduces subjectivity, potentially skewing interpretations. The study also lacks exploration of advanced *machine learning* methods, which could offer more nuanced analysis despite their computational demands. Future research should consider these limitations for broader applicability and improved understanding.

The study stands out for its comprehensive approach and interdisciplinary impact, merging traditional literary analysis with computational techniques to enhance understanding of similes in Arabic rhetoric. Its strengths are in its innovative methodology and the fresh insights it offers into the structure and function of similes. To build upon these strengths, future research should consider a more diverse corpus and potentially incorporate advanced computational methods to further refine the analysis of similes in Arabic rhetoric. This will not only address the current limitations but also broaden the

scope and depth of understanding in this important area of study.

However, there are weaknesses to consider. The study's reliance on rule-based methods for simile extraction, while practical and less resource-intensive, may not capture the nuanced understanding that advanced *machine learning* methods offer. This could limit the ability to adapt to the varied and complex nature of similes in Arabic texts. Additionally, the manual review process, a crucial step in validating the output of the algorithm, introduces an element of subjectivity (Gurus, 2021). This subjectivity could potentially affect the accuracy and objectivity of simile classification, as the personal interpretations and biases of the reviewers might influence the final analysis.

The study contributes substantially by refining simile detection methodologies in NLP, adding depth to literary interpretations in Arabic criticism, and advancing cognitive linguistic theories on metaphor and simile processing. By providing a more detailed understanding of similes, the research enhances the appreciation of stylistic choices and thematic depth in these texts. This new lens allows for a richer, more nuanced engagement with Arabic literature, potentially leading to more profound interpretations and discussions within the literary community.

In the realm of cognitive linguistics, the study significantly contributes to the understanding of how the human mind processes figurative language, with a particular focus on similes. This insight is vital for comprehending the cognitive mechanisms underpinning language comprehension and emotional responses to literature. It broadens the scope of research in cognitive linguistics, offering valuable data on the interaction between language and thought.

Moreover, in the field of natural language processing, the research provides crucial data and methodologies for the development of algorithms capable of detecting and interpreting similes in Arabic texts. This contribution is particularly valuable for advancing computational linguistics, as it addresses the complex challenge of processing figurative language—a task that machines have traditionally struggled with. By improving these capabilities, the study aids in the creation of more sophisticated language processing tools, enhancing the interaction between humans and technology in language understanding.

The study of Arabic rhetoric's similes has limitations, including a singular corpus like *Al-Abrat* and potential algorithmic bias. It suggests incorporating diverse texts for a more representative understanding. The manual review process introduces subjectivity, potentially skewing interpretations. The study also lacks exploration of advanced *machine learning* methods, which could offer

more nuanced analysis despite their computational demands. Future research should consider these limitations for broader applicability and improved understanding.

The study stands out for its comprehensive approach and interdisciplinary impact, merging traditional literary analysis with computational techniques to enhance understanding of similes in Arabic rhetoric. Its strengths are in its innovative methodology and the fresh insights it offers into the structure and function of similes. To build upon these strengths, future research should consider a more diverse corpus and potentially incorporate advanced computational methods to further refine the analysis of similes in Arabic rhetoric. This will not only address the current limitations but also broaden the scope and depth of understanding in this important area of study.

In conclusion, the findings of this study support both research hypotheses:

- *Hypothesis 1:* The Python-based approach successfully detected and classified similes with high accuracy, as demonstrated by the identification of 212 similes, 130 of which were confirmed through manual validation.
- *Hypothesis 2:* The classification of similes based on Abdul-Raof's framework revealed distinct patterns in simile usage, with perceptible-perceptible similes being the most frequent. This aligns with the cognitive and rhetorical functions of similes in classical Arabic literature, providing deeper insights into their use in Arabic texts.

These results underscore the effectiveness of combining computational techniques with manual oversight.

Study Limitation

This study, while groundbreaking, has several limitations that warrant attention. Its reliance on a single corpus, *Al-Abrat*, restricts the generalizability of findings across Arabic literature. The rule-based algorithm, though effective, struggles with nuanced or unconventional similes, necessitating manual validation that introduces subjectivity. Additionally, the study focuses solely on similes, excluding other rhetorical devices, and does not account for regional dialects or linguistic variations within Arabic. The computational methods used, while innovative, are resource-intensive and may benefit from integration with advanced *machine learning* techniques. Future research could address these limitations by expanding the corpus, incorporating diverse rhetorical devices, and utilizing more sophisticated computational approaches to enhance scalability and accuracy.

Impact

Studying has a significant impact across multiple disciplines. In literary criticism, it introduces a novel methodology for analyzing figurative language, thereby enriching the interpretation of Arabic texts. From a cognitive linguistics perspective, the research offers valuable insights into how similes shape cognitive processing, enhancing our understanding of figurative language comprehension. Furthermore, in the field of natural language processing, the development of simile detection algorithms represents a notable advancement, with practical applications in machine translation and sentiment analysis.

Conclusion

Mustafa Lutfi al-Manfaluti's *Al-'Abrat* ("The Tears") is a seminal work in modern Arabic literature, exploring themes of love, sacrifice, and hardship through poignant storytelling. Its only English translation, *The Tears* (1915), by Majid Khan Malik Saddiqui, offers 10 evocative tales, bridging linguistic and cultural gaps while highlighting universal themes of emotional and cultural depth.

The study focuses on similes as a rhetorical device in Arabic literature, particularly in *Al-'Abrat*, using Abdul-Raof's (2006) classification framework. It integrates computational techniques, such as a Python-based algorithm, to detect and classify 212 similes into categories like perceptible-perceptible and cognitive-cognitive. This approach combines traditional literary criticism with modern natural language processing (NLP), addressing gaps in figurative language analysis, particularly similes, in Arabic rhetoric.

Findings underscore similes' cognitive and rhetorical significance, revealing patterns that enhance emotional engagement and cultural understanding. While effective, the study notes limitations, including reliance on a single corpus, subjectivity in manual validation, and the need for advanced *machine learning* methods for nuanced simile detection.

The research makes interdisciplinary contributions to literary criticism, cognitive linguistics, and NLP, advancing methodologies for simile analysis and enriching understanding of Arabic literature's rhetorical depth. Future studies are encouraged to broaden the corpus, incorporate diverse rhetorical devices, and leverage sophisticated computational tools for enhanced analysis.

Acknowledgments

The authors extend their profound gratitude to all individuals who provided invaluable assistance during the data acquisition phase of this study. Further appreciation is directed towards

the administrative and academic leadership of University Putra Malaysia (UPM) for their unwavering support and facilitation, which significantly contributed to the successful realization of this research endeavor.

ORCID iDs

Rasha Talal Al Zahrawi  <https://orcid.org/0009-0003-0714-7048>

Syed Nurulakla Syed Abdullah  <https://orcid.org/0000-0002-4444-8822>

Muhammad Alif Redzuan Abdullah  <https://orcid.org/0000-0002-8684-0769>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data supporting the findings of this study consist of textual materials and translation outputs that are not publicly archived due to copyright and privacy considerations. These data are available from the corresponding author on reasonable request.

Supplemental Material

Supplemental Material for this article is available online.

References

- Abdul-Raof, H. (2006). *Arabic rhetoric: A pragmatic analysis*. Routledge.
- Abugharsa, A. (2022). *Metaphor detection in poems in Misurata Arabic subdialect: An LSTM model* [Master's thesis]. Montclair State University. <https://digitalcommons.montclair.edu/etd/1080>
- Adeyemi, D. K. (2017). The essence of al-Balāghah (rhetoric) to Arabic and Yoruba languages. *IOSR Journal of Humanities and Social Science*, 22(10, Ver. 9), 27–31. <https://doi.org/10.9790/0837-2210092731>
- Admin. (2019, March 7). Figurative language—Examples and definition. *Literary Devices*. https://literarydevices.net/figurative-language/#google_vignette
- Al-Ghamdi, D., Abdul-Mageed, M., Diab, M., & El-Barbary, A. (2020). A survey of Arabic natural language processing: Challenges, applications, and opportunities. *Computational Linguistics*, 46(2), 1–52. https://doi.org/10.1162/coli_a_00368
- Al-Manfaluti, M. L. (1916). Al—‘Abrat [The tears]. *Hindawi Foundation*. <https://www.hindawi.org/books/29373683/>
- Al-Manfaluti, M. L. (n.d.) *The tears* (the English Translation of Al-Abarat: Mustafa Lutfi Al-Manfaluti (M. K. M. Saddiqui, Trans.). Amazon.com: Books.
- Al-Mubarak, R., & Hill, I. E. J. (2023). A descriptive study of Arabic rhetorics. In *Routledge handbooks* (pp. 148–162). Routledge. <https://doi.org/10.4324/9781003195276-13>
- Alshammari, J. N. M. (2016). Analyzing Arabic translation methods of English similes: A case study of *the Old Man and the Sea* by Ernest Hemingway. *Theory and Practice in Language Studies*, 6(3), 485–494. <https://doi.org/10.17507/tpls.0603.05>
- Al-Sharafi, A. G. M. (2004). *Textual metonymy: A semiotic approach*. Palgrave Macmillan.
- Al Zahrawi, R. T., Syed Abdullah, S. N., Brahimi, T., Abdullah, M. A. R., & Mustapha, N. F. (2024). Bibliometric analysis of Arabic Rhetoric in the translation and transcreation of literary texts. *Cogent Arts & Humanities*, 11(1), Article 2428483. <https://doi.org/10.1080/23311983.2024.2428483>
- Balagha, A. R. (2020). *Predication in Arabic rhetoric*. <https://www.learnarabiconline.com/predication/>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Edunov, S., Ott, M., Ranzato, M., & Auli, M. (2020, July). *On the evaluation of machine translation systems trained with Back-translation* [Conference session]. 58th Annual Meeting of the Association for Computational Linguistics (pp. 2836–2846). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.253>
- Elgabry, H., Attia, S., Abdel-Rahman, A., Abdel-Ate, A., & Girgis, S. (2021). *A contextual word embedding for Arabic Sarcasm detection using data augmentation and neural models*. Proceedings of the Sixth Arabic Natural Language Processing Workshop (WANLP 2021).
- Ellis, M. (2024, October 24). Figurative language examples: 6 common types and definitions. *Grammarly*. <https://www.grammarly.com/blog/writing-tips/figurative-language/>
- Fass, D. (1991). “met*”: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1), 49–90.
- Feldman, J. A. (2004, August 23–27). *Computational cognitive linguistics* [Conference session]. 20th International Conference on Computational Linguistics (COLING 2004), Geneva, Switzerland (pp. 1114-es). Association for Computational Linguistics. <https://doi.org/10.3115/1220355.1220515>
- Fu, X. (2022). Metaphor recognition of English learners based on machine learning algorithm. *Wireless Communications and Mobile Computing*, 2022(1), 1–11. <https://doi.org/10.1155/2022/1177896>
- Górniak-Woźniak, A. (2023, July 18). Back translation—Is control the highest form of trust? *Alingua Translation Agency*. *Alingua Translation Agency*. <https://alingua.pl/en/back-translation-is-control-the-highest-form-of-trust/>
- Grafiati. (2021, June 4). Arabic rhetoric. *Grafiati blog*. <https://www.grafiati.com/en/literature-selections/arabic-rhetoric/journal/>

- Gurus, C. (2021). *Using manual review properly. Chargeback Gurus: Smart, trusted, tech-enabled*. <https://www.chargebackgurus.com/blog/manual-review>
- Heaton, S., Teece, D., & Agronin, E. (2023). Dynamic capabilities and governance: An empirical investigation of financial performance of the higher education sector. *Strategic Management Journal*, 44(2), 520–548. <https://doi.org/10.1002/smj.3444>
- Humoud. (2020). Detecting Arabic characters with regex [GitHub Gist]. *GitHub*. <https://gist.github.com/Humoud/f40f58cd85c5935a444c>
- Hussein, A. A., & Kuflik, T. (2023). Automatic identification of rhetorical elements in classical Arabic poetry. *Digital Humanities Quarterly*, 17(2). <https://doi.org/10.17613/000673>
- Israel, M., Harding, J. R., & Tobin, V. (2004). On simile. In M. Achard & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 123–135). CSLI Publications.
- Jimi, J., & Shimada, K. (2024). Simile identification with pseudo data acquisition and re-labeling. *International Journal of Smart Computing and Artificial Intelligence*, 8(1), 1–14. <https://doi.org/10.52731/ijscai.v8.i1.780>
- Kindree, V., & Barksdale, P. J. (2022, September 10). Simile, metaphor, and symbol: Figurative language boot camp. *AuthorToolboxBlogHop. Words like trees*. <https://wordsliketrees.wordpress.com/2019/10/13/simile-metaphor-symbol-fig-lang-1/>
- Kızıklı, Z. (2007). Edebî bir sanat olarak arap Belâgatinde teşbih [Simile as a literary art in Arabic rhetoric]. *Dini Araştırmalar*, 9(27), 223–246.
- Lai, H., & Nissim, M. (2024). A survey on automatic generation of figurative language: From rule-based systems to large language models. *ACM Computing Surveys*, 56(10), 1–34. <https://doi.org/10.1145/3654795>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar, I. Theoretical prerequisites*. Stanford University Press.
- Literary Devices. (2020, September 15). *Simile—Examples and definition of simile*. <https://literarydevices.net/simile/>
- Liu, Y., Hu, X., & Liu, J. (2022). A corpus-based study on Chinese and American students' rhetorical moves and stance features in dissertation abstracts. *Frontiers in Psychology*, 13, Article 1004744. <https://doi.org/10.3389/fpsyg.2022.1004744>
- Mahmoud, A., & Zrigui, M. (2017). *Semantic similarity analysis for paraphrase identification in Arabic texts*. Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation (PACLIC 31), pp. 404–413.
- Mason, Z. J. (2004). CorMet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1), 23–44. <https://doi.org/10.1162/089120104773633376>
- Mirela. (2024, June 7). Back translation: What it is, benefits, limitations. *POEditor blog*. <https://poeditor.com/blog/back-translation/>
- Moreau, F. (2003). *Al-Balāgha. Al-Madkhal li-dirāsāt al-suwar al-bayāniyya* (M. Al-Wahī, & J. Aisha, Trans.). Afriqyā l-Sharq.
- Ortony, G. A., & Miller, G. A. (1993). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (pp. 357–400). Cambridge University Press. <https://doi.org/10.1017/CBO9781139173865.019>
- Rahman, R. A., Jusoh, T. G. T., Abdullah, M. N., & Zin, H. M. (2013). Research on the writing history of Arabic rhetoric studies and its importance. *Academic Journal of Interdisciplinary Studies*, 2(9), 517. <https://doi.org/10.5901/ajis.2013.v2n9p517>
- Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020, November). *COMET: A neural framework for MT evaluation* [Conference session]. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP), 2020 (pp. 2685–2702).
- Ren, D., Zhang, P., Li, Q., Tao, X., Chen, J., & Cai, Y. (2020). A hybrid representation-based simile component extraction. *Neural Computing & Application*, 32, 14655. <https://doi.org/10.1007/s00521-020-04818-6>
- Ryan, E. (2023, August 12). What is a simile? | Meaning, definition & examples. *Scribbr blog*. <https://www.scribbr.com/rhetoric/simile/>
- Samson, V. A. (2019, October 21). *Fun with back translation*. P. B.R. Tradiling. <https://www.tradiling.net/fun-with-reverse-translation/>
- Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, 41(4), 579–623.
- Veale, T., & Hao, Y. (2007). Learning to understand figurative language: From similes to metaphors to irony. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29, 683–688.
- von Grunebaum, G. E. (2012). Balāgha. In P. Bearman, T. Bianquis, C. E. Bosworth, E. van Donzel, & W. P. Heinrichs (Eds.), *Encyclopaedia of Islam* (2nd ed.). Brill. https://referenceworks.brillonline.com/entries/encyclopaedia-of-islam-2/balagha-COM_0075
- Whatley, S. (2024, May 22). Forward translation and back translation: Do they improve quality? | Comms multilingual. *Comms Multilingual*. <https://www.commsmultilingual.com/comms-news-blog/do-forward-translation-and-back-translation-improve-translation-quality/>
- Yaseena, M., Ibrahimc, H., & Sayyeddb, S. (2022). Modes of persuasion for women in English and Arabic advertisements. *European Journal of Applied Linguistics and Literature*, 8, 180–190. <https://doi.org/10.32601/ejal.911552>
- Zafer, K. (2007). Simile in the Arabic rhetoric as a literary figure: This article deals with the concept of simile in the Arabic language within the context of rhetoric. *Dini Araştırmalar*, 9, 223–246.
- Ziheng, Z., & Bhat, S. (2021). Idiomatic expression identification using semantic compatibility. *Transactions of the Association for Computational Linguistics*, 9, 1546–1562.