Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping gender, agency, and objectification in Arabic poetry: an ontology-guided computational study

Anoud A. Alhamad
Princess Nourah bint Abdulrahman University
Wadee A. Nashir
Digital Dhad Experts
Hassan Alshamrani
King Saud University
Badriyya B. Al-onazi
Princess Nourah bint Abdulrahman University
Reema G. Al-anazi
rgalanizy@pnu.edu.sa
Princess Nourah bint Abdulrahman University
Wala F. Bin subait
Princess Nourah bint Abdulrahman University
Najla I. Al-shathry
Princess Nourah bint Abdulrahman University
Azizah M. Almutairi
Princess Nourah bint Abdulrahman University

Article

Keywords:

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11178991/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Mapping gender, agency, and objectification in Arabic poetry: an ontology-guided computational study

Anoud A. Alhamad[1], Wadee A. Nashir[2], Hassan Alshamrani[3,*], Badriyya B. Al-onazi[4], Reema G. Al-anazi[5], Wala F. Bin subait[4], Najla I. Al-shathry[4] Azizah M. Almutairi[4]

[1] Department of English Literature, College of Languages, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[2] Digital Dhad Experts, Riyadh, Saudi Arabia

[3] Department of Arabic Language Studies as a Second Language, College of Languages and Sciences, King Saud University, Riyadh, Saudi Arabia

[4] Department of Language Preparation, Arabic Language Teaching Institute, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[5] Department of Arabic Language and Literature, College of Humanities and Social Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[*] Corresponding author: rgalanizy@pnu.edu.sa

Abstract. Gender, agency, and objectification are important dimensions of how participants and relationships are represented in poetry, yet their systematic study in Arabic verse remains difficult because meaning is often conveyed through metaphor, ellipsis, compressed syntax, and context-dependent relations. This study investigates how gendered agency and objectification are represented across a large corpus of Arabic poetry and examines whether computational analysis can capture these literary-semantic relations while preserving textual evidence and interpretive uncertainty. A neuro-symbolic computational framework was developed to represent participants, events, agency, gender representation, and objectification as explicit, evidence-linked semantic relations. Large language model extraction was combined with formal constraint validation and expert adjudication. Evaluated against an expert-annotated sample of 4,500 verses, automated semantic extraction achieved an F1-score of 0.884 on the hold-out test set. Following evaluation and targeted adjudication within the annotated sample, the resulting constraint-guided pipeline was applied across 31,209 single-verse poems spanning twelve historical periods and eight poetic genres. Corpus-level analysis identified distinct distributions in gendered agency: male-coded participants were classified as active in 61.8% of agency-related instances, compared with 48.2% for female-coded participants, while the latter showed higher proportions of reactive and passive involvement. Objectification was distributed across multiple forms, with symbolic objectification emerging as the most frequent category (38.4%), exceeding direct physical objectification (21.6%). These results make recurrent patterns of gendered literary representation systematically inspectable at corpus scale while retaining links to textual evidence, ambiguity, and provenance. Ultimately, the study demonstrates how evidence-grounded computational methods can support large-scale investigation of literary representation while retaining interpretive boundaries.

Keywords: Arabic poetry; gender representation; literary agency; objectification; digital humanities; computational literary studies; literary semantics; neuro-symbolic AI.

---

1 Introduction

Understanding how gender, agency, and objectification are represented in literary texts remains a significant challenge for computational literary studies. In cultural heritage and digital humanities, computational methods and knowledge graphs can represent fragmented textual evidence as interconnected conceptual structures (Haslhofer et al., 2018; Huang et al., 2023). Large Language Models (LLMs) can support such analysis through open-ended text comprehension, but their use alone is problematic for highly implicit and figurative discourse. Generative models lack intrinsic formal governance and require domain guidance and output validation (Schilling-Wilhelmi et al., 2025). In figurative discourse, they may introduce unsupported source domains or misidentify metaphorical mappings (Tong et al., 2024; Wachowiak and Gromann, 2023). They can also generate factual and logical inconsistencies; therefore, computational literary analysis requires representations that preserve textual evidence while making extracted interpretations structurally verifiable (Huang et al., 2025).

This challenge motivates a neuro-symbolic approach combining semantic modelling, logical constraints, and explicit uncertainty representation. Prior ontology-based information extraction frameworks (Zaman et al., 2021) and formally constrained generation methods (Geng et al., 2023) indicate that grounding probabilistic language models in structural rules can improve the validity of extracted information. However, existing approaches often treat classification, semantic role labelling (SRL), and information extraction as separate tasks, leaving a need for integrated representations that capture implicit relational meanings in literary texts while retaining their evidential and interpretive provenance.

Classical Arabic poetry provides a particularly demanding setting for this problem and a rich domain for examining how gendered relations and agency are encoded through explicit and implicit expression. Arabic natural language processing (NLP) faces challenges arising from complex nonconcatenative morphology, orthographic ambiguity, and limited resources for historical and literary domains (Darwish et al., 2021; Farghaly and Shaalan, 2009). As a pro-drop language, Arabic frequently omits subject pronouns, complicating zero-pronoun and coreference resolution (Aloraini et al., 2025). Classical Arabic poetry further relies extensively on indirect and rhetorically dense expression (Abd Alhadi et al., 2023). Consequently, identifying who acts, who is acted upon, and how relations such as objectification or gaze are constructed requires interpretation across linguistic context, situational factors, and cultural knowledge rather than surface lexical matching alone (Zibin et al., 2025). Conventional NLP tasks such as sentiment analysis and named entity recognition provide useful foundations but generally operate at broader document-, word-, or entity-span levels (Abo et al., 2019; El Moussaoui and Loqman, 2024), making them insufficient on their own for modelling fine-grained intra-textual literary semantics.

Recent computational studies have produced large-scale Arabic poetry corpora (Al-Onazi et al., 2025) and models for thematic classification (Al-Anazi et al., 2025). Nevertheless, much of this work focuses on document-level tasks, including authorship attribution and poet identification (Shoubaki et al., 2024), rather than semantic relations within individual poetic texts. This gap is particularly relevant to gender and agency. Computational distant reading has recently demonstrated that gender ideology is deeply legible in the distributional clustering of body-part nouns and the action verbs predicated of male versus female characters within English novelistic corpora (Chen et al., 2024). However, in Arabic computational research, gender has largely been restricted to a variable for authorship profiling (Alsmearat et al., 2017), while agency has been examined as a broader discursive pattern, such as in political texts (Karimullah, 2020). The present study extends the logic of distributional gender analysis to a highly figurative, diachronic Arabic poetic corpus, directly representing gendered agency, objectification, and gaze as structured, intra-textual semantic relations.

This study addresses this gap by treating Classical Arabic poetry as a humanistic inquiry supported by controlled semantic representation. Rather than positioning computation as an end in itself, the framework transforms implicit and figurative textual evidence into auditable representations that support corpus-scale distant reading of gendered relations. Its contribution thus centres on systematic literary analysis, with semantic infrastructure serving as the methodological means of making these relations explicit, traceable, and consistently analysable. To operationalise this inquiry, the study is guided by three core research questions (RQs):

- **RQ1:** How can implicit gendered relations, such as agency and objectification, be computationally represented and extracted from highly figurative Classical Arabic poetry without suppressing inherent literary ambiguity?
- **RQ2:** To what extent does integrating formal ontological constraints with probabilistic language models improve the structural consistency and evidence traceability of literary-semantic extraction, while distinguishing formal validity from interpretive validity?
- **RQ3:** What corpus-level patterns of gendered agency and objectification emerge when distant reading is applied to a large-scale, diachronic corpus of Arabic verse?

The contributions of this study to digital humanities and computational literary studies are:

1. A **corpus-level analysis of gendered literary representation** in Classical Arabic poetry, identifying patterned differences in agency and objectification across a large, diachronic corpus while retaining links to the underlying textual evidence.
2. An **evidence-grounded operationalisation** of agency, objectification, and gaze as interconnected literary-semantic relations, allowing these interpretive constructs to be examined systematically without treating their computational representation as equivalent to literary meaning.
3. A **controlled semantic methodology** combining contextual extraction, explicit ontological constraints, uncertainty preservation, and expert adjudication to make such literary relations traceable and reproducibly analysable at corpus scale.
4. An empirical demonstration that computational consistency and literary interpretation are distinct: formal validation improves structural admissibility, whereas context-dependent cases still require expert interpretation.

2 Related Work

Computational study of gender representation in literature brings together literary interpretation, Arabic literary studies, computational literary studies, and natural language processing. Literary scholarship provides concepts for gendered agency, objectification, and relations of looking, while computational methods enable analysis across larger collections. The central challenge is to represent these context-dependent phenomena explicitly and with evidential grounding without reducing them to undifferentiated labels. This review considers literary scholarship, Arabic computational studies, distant reading, and semantic modelling.

2.1 Gender, Agency, and Objectification in Literary Studies

Gender in literary analysis concerns not only authors or characters but how participants are positioned within actions, relationships, and representational structures. Agency is particularly relevant because grammatical subjecthood does not necessarily correspond to narrative or causal control (Stuhler, 2024). Objectification complements this perspective by examining how a represented person is constituted through another participant's perception, desire, or evaluative framing, while the male gaze concerns the directional relation through which such perception is organized (Luo et al., 2024). These concepts are therefore treated here as relational literary constructs rather than lexical categories: agency concerns a participant's role in producing or responding to events, objectification concerns the representational positioning of that participant, and gaze concerns the directionality of perception. Their computational operationalisation consequently provides an analytical representation of these constructs, not a claim that the representation exhausts their literary meaning.

2.2 Gender Representation and Computational Analysis of Arabic Poetry

Arabic poetry provides a distinctive setting because gendered representation is embedded in conventionalized poetic forms as well as description, address, remembrance, and interaction. Speakers and represented participants may occupy positions as observers, agents, respondents, or objects of representation, making gender representation a question of relational semantics rather than simple frequency or demographic profiling.

Arabic NLP faces challenges from morphology, orthographic variation, resource diversity, and non-standard domains (Darwish et al., 2021; Farghaly and Shaalan, 2009). Classical and rhetorical Arabic further relies on ellipsis, indirect expression, and metaphorical substitution, which may leave relevant participants or relations implicit (Abd Alhadi et al., 2023; Zibin et al., 2025); zero-pronoun and coreference phenomena add difficulty when arguments are omitted from the surface form (Aloraini et al., 2025). Recent computational studies provide foundations for Arabic literary analysis through resources such as the Diwan corpus (Al-Onazi et al., 2025) and work on meter, poet or era attribution, thematic and rhetorical classification (Mutawa and Alrumaih, 2025; Shoubaki et al., 2024; Al-Anazi et al., 2025; Nashir et al., 2026b). Gender has also been examined through authorship profiling (Alsmearat et al., 2017), and agency as a broader discourse phenomenon in Arabic political corpora (Karimullah, 2020). These studies establish important foundations but generally operate at different analytical levels; the reviewed literature does not directly integrate gendered agency, objectification, and gaze as interconnected relations within individual poetic events.

2.3 Computational Literary Studies and Distant Reading

Computational literary studies and distant reading enable investigation of distributions, recurrent patterns, networks, and changes across large collections. Knowledge graphs have likewise been used in digital humanities and cultural heritage to connect heterogeneous textual and bibliographic information through structured representations (Haslhofer et al., 2018; Huang et al., 2023; Stranisci et al., 2023), while narrative and literary ontologies represent events, characters, and temporal or relational structures in machine-actionable form (Meghini et al., 2021; Kalathil et al., 2024).

For Arabic poetry, computational resources support corpus-scale analysis through meter recognition, authorship attribution, language modelling, and thematic classification (Mutawa and Alrumaih, 2025; Shoubaki et al., 2024; Qarah, 2024; Al-Anazi et al., 2025). Their analytical units are often documents, verses, authors, themes, or predefined categories, supporting classification and distributional questions but providing less direct access to internal semantic relations connecting participants to actions, observers, or objectifying descriptions. Distant reading thus supplies scale but not necessarily the semantic granularity required to investigate agency and objectification.

2.4 Semantic Role labelling, LLM Extraction, and Structured Representation

Semantic Role labelling (SRL) provides an established basis for identifying predicates and participants. Arabic resources including Camel Treebank (Habash et al., 2022) and recent multi-layered resources for Classical Arabic (Nashir et al., 2025a) support recovery of event participants and structural relations. However, conventional roles such as Agent, Patient, and Theme do not exhaust the distinctions required here: participants may initiate events, respond to others, undergo events without meaningful control, or remain ambiguous, while objectification and gaze involve relations that cannot be reduced to a single predicate–argument label. Literary and narrative processing highlights the difficulties of complex discourse (Sims et al., 2019), and Arabic adds ellipsis, flexible word order, and omitted arguments (Nashir et al., 2025b). SRL therefore provides a foundation without supplying the full literary-semantic vocabulary required for gendered agency, objectification, and gaze.

LLMs complement SRL by extracting contextual relations from complex texts. Domain-specific information extraction demonstrates their use for deriving structured relations (Arsenyan et al., 2024), while work on machine-actionable resources emphasizes structural guidance for generated outputs (Schilling-Wilhelmi et al., 2025). This is especially relevant to literary language, where metaphor and context can permit multiple plausible readings. Metaphor studies report difficulties in recovering appropriate mappings (Tong et al., 2024; Wachowiak and Gromann, 2023), and broader work identifies factual hallucination as a persistent concern (Huang et al., 2025). Crucially, computational studies of Arabic rhetoric (Al-Anazi et al., 2026a) and recent evaluations (Zibin et al., 2025) illustrate that purely neural models continue to misinterpret figurative Classical Arabic poetry due to cultural constraints rather than model scale. Our ontology-guided pipeline responds directly to this gap by externalising interpretive categories into an explicit, inspectable schema.

These limitations motivate structured semantic representations and explicit constraints alongside probabilistic language processing. Ontology-based extraction defines admissible concepts and relations (Zaman et al., 2021);

grammar- and schema-constrained generation and related approaches seek to improve output validity or reduce inconsistent relations (Geng et al., 2023; Kuculo et al., 2025; Shiri et al., 2024; Zhou et al., 2026); and constraint-based semantic modelling has been applied to Classical Arabic rhetorical analysis (Al-Anazi et al., 2026b). For computational literary studies, however, structural validity is only one requirement. Representations must retain event and participant relations, evidence-supported gender inference, differentiated agency states, forms of objectification, directed gaze, and uncertainty where multiple interpretations remain plausible. Linking each assertion to its textual evidence is likewise essential. Ontology, constraint validation, uncertainty representation, expert adjudication, and knowledge-graph construction therefore make corpus-scale literary analysis inspectable and reproducible rather than constituting independent technical objectives.

2.5 Research Gap

The literature provides complementary capabilities: literary scholarship supplies concepts for gendered agency, objectification, and gaze; Arabic literary and computational research demonstrates corpus-scale poetry analysis; distant reading provides comparative scale; and NLP, SRL, ontology-based extraction, and constrained LLM methods provide structured means of representing semantic information. The remaining gap is their systematic integration for the analytical problem addressed here.

Existing computational analysis of Arabic poetry remains largely focused on document-, author-, thematic-, or formal-level properties (Al-Anazi et al., 2025; Qarah, 2024; Shoubaki et al., 2024). Computational treatments of gender and agency show that these concepts can be operationalised at scale (Stuhler, 2024; Luo et al., 2024), but relevant approaches have largely been developed in other literary traditions or at different analytical levels. Conventional SRL and narrative representations provide event structures without directly encoding the multidimensional literary constructs required here, while LLMs provide contextual inference but require structural governance when their outputs become scholarly data (Huang et al., 2025; Schilling-Wilhelmi et al., 2025).

The prevailing research gap is therefore not an absence of computational techniques, but rather the limited systematic integration of these distinct strands to model gender, agency, objectification, and gaze as interconnected, intra-textual relations within Classical Arabic poetry. As synthesised in Table 1, MARAYA addresses this methodological void by introducing an event-centred semantic representation. This approach combines literary-critical categories, contextual extraction, deterministic constraint validation, explicit uncertainty preservation, and expert adjudication into an auditable knowledge graph. The ultimate objective is to facilitate corpus-scale literary inquiry that firmly anchors computational interpretations to their textual provenance, explicitly rejecting the premise that generated classifications can serve as self-sufficient literary conclusions.

3 Epistemic Foundations and Interpretive Principles

The computational analysis of gendered agency and objectification in Classical Arabic verse requires an approach that mediates between distant reading and the compressed, figurative, and context-dependent character of poetic discourse. Meaning may emerge through metaphor, ellipsis, implicit reference, and rhetorical indirection, making it difficult to map literary constructs directly onto observable lexical features. Treating such constructs as unambiguous categories risks reducing complex interpretive phenomena to simplified computational labels. The aim is therefore not to automate an objective literary "truth" or replace close reading, but to provide an auditable means of examining relational patterns across a corpus while acknowledging the historical distance and interpretive plurality of the texts.

This approach requires computational claims to remain traceable to inspectable textual evidence (Kuhn, 2019). Rather than treating agency, objectification, and related constructs as isolated labels, the framework represents them through relations among participants, events, and their textual provenance (Ciotti, 2016). Such traceability does not imply that textual evidence always determines a single interpretation. Metaphorical language, implicit reference, and contextual ambiguity may sustain more than one plausible reading; these possibilities should not be collapsed automatically into categorical decisions. Uncertainty is therefore treated as an epistemic consideration in the interpretation of computational outputs, distinguishing claims supported by textual evidence from those requiring further scholarly judgement (Chen et al., 2019).

Table 1: Position of MARAYA within Related Computational and Literary Approaches

Approach / Research Tradition	Primary Semantic Unit	Gender / Agency	Ontology / Schema	Constraint / Validation	Uncertainty / Evidence	Typical Scope
Standard Arabic NLP (Aladeemy et al., 2024; Qu et al., 2024; Shaalan, 2014)	Sentence / Entity	Not primary	Task-specific	Task-dependent	Task-dependent	Prose / News
Arabic Poetry Analytics (Al-Anazi et al., 2025; Qarah, 2024; Shoubaki et al., 2024)	Verse / Document	Limited / indirect	Task-specific	Task-dependent	Varies	Arabic Poetry
Literary SRL (Sims et al., 2019)	Predicate–argument	Not primary	Role inventory	Not central	Not central	Literary Texts
Narrative and Literary KGs (Kalathil et al., 2024; Meghini et al., 2021; Stranisci et al., 2023)	Event / Entity	Not primary	Explicit	Varies	Varies	Narrative / Cultural Data
LLM-based Extraction (Arsenyan et al., 2024; Huang et al., 2025; Schilling-Wilhelmi et al., 2025)	Contextual relation	Task-dependent	Prompt / Schema	Often external	Evidence-dependent	Domain-specific / Open Domain
Constraint-guided Semantic modelling (Zaman et al., 2021; Geng et al., 2023; Kuculo et al., 2025; Al-Anazi et al., 2026b)	Structured relation	Task-dependent	Explicit	Explicit	Varies	Scientific / Technical / Rhetorical
MARAYA	**Event–Participant Relation**	**Explicit**	**Literary-Semantic**	**Deterministic**	**Explicit**	**Classical Arabic Poetry**

A further distinction is necessary between formal semantic admissibility and hermeneutic validity. Probabilistic language models can propose contextually plausible interpretations, but plausibility alone does not establish that an assertion is textually warranted or structurally consistent. The framework therefore separates candidate extraction from the checking of predefined ontological constraints (Guarino, 1995). Such checks assess conformity with specified conditions; they do not establish literary truth or resolve every semantic contradiction. Interpretations that remain ambiguous or require contextual clarification must consequently remain open to human scholarly judgement. This distinction positions the computational representation as an instrument for critical literary inquiry rather than a substitute for it (see Supplementary Note S1 and Supplementary Fig. S1 for the extended epistemic rationale and verification procedures).

4 Operationalising Literary Relations: An Event-Centric Semantics

The analysis of gendered agency, objectification, and gaze across Classical Arabic poetry requires literary-critical concepts to be represented as explicit relations while preserving their dependence on textual context. These phenomena cannot be reduced to isolated lexical indicators: agency is not equivalent to grammatical subjecthood, objectification may be conveyed through physical description or symbolic and figurative representation, and gaze concerns the direction of represented perception rather than the mere occurrence of visual language. MARAYA therefore operationalises these constructs within an event-centred semantic representation designed to support corpus-scale comparison without treating computational categories as exhaustive accounts of literary meaning.

4.1 Event-Centred Modelling of Classical Arabic Verse

The poetic event provides the principal context for representing actions, participants, and their semantic relations, following event-oriented approaches to knowledge representation (Brown et al., 2017; Meghini et al., 2021). MARAYA distinguishes a normalised action from its occurrence in a particular poetic context: the same action may involve different participants or acquire different implications through metaphor, rhetorical address, or surrounding relations. This distinction is especially relevant to Classical Arabic verse, where syntactic compression, ellipsis, and zero-anaphora may leave participants or their roles implicit. The event-centred representation makes these

relations available for systematic examination while retaining the need to assess inferred participants against the textual evidence.

The hierarchical organisation of the ontology—connecting overarching literary constructs (Gender, Agency, Objectification, and Gaze) to contextualised poetic events, participant roles, and textual provenance—is formally specified and illustrated in Supplementary Note S2 and Supplementary Fig. S2.

4.2 Operational Definitions of Gender, Agency, Objectification, and Gaze

Gender is represented as an epistemically qualified attribute rather than assumed to be explicit in every verse. The categories Explicit, Inferred, and Unknown distinguish directly indicated attribution from contextual inference and cases where the available evidence does not support a grounded assignment (Ehrmann et al., 2023). Gender attribution is kept analytically distinct from semantic-role assignment: uncertainty about a participant's gender does not automatically prevent the analysis of otherwise supportable relations involving that participant.

Agency is treated as a participant's relational involvement in an event, with attention to causal contribution and meaningful control rather than grammatical subjecthood alone (Stuhler, 2024). MARAYA distinguishes Active agency, in which a participant initiates an action; Reactive agency, in which a participant responds to an action; Passive involvement, in which a participant is affected without meaningful control; and Ambiguous cases that permit competing readings. These categories provide an operational vocabulary for comparison, not an exhaustive theory of literary agency.

Objectification is differentiated into Physical, Emotional, Symbolic, and Mixed forms. This distinction allows corporeal, affective, and symbolic representations to be examined separately rather than treating objectification as a single undifferentiated phenomenon. Gaze is represented directionally through Male-to-Female, Female-to-Male, Mutual, and Self-directed relations between a perceiver and a perceived entity (Luo et al., 2024). These categories describe the study's analytical operationalisation; they do not, by themselves, establish authorial intention or the lived experience of historical populations.

Because metaphor, ellipsis, and contextual dependence may support more than one plausible interpretation, a formally admissible assertion is not necessarily a definitive literary reading. MARAYA therefore distinguishes semantic proposal, structural verification, and unresolved interpretation. The representation supports systematic comparison while retaining links to textual evidence and leaving context-dependent cases open to scholarly scrutiny; the formal uncertainty representation and validation requirements are detailed in Supplementary Note S2.

5 Extraction Workflow: Controlled Inference and Expert Adjudication

MARAYA translates the event-centred representation described in Section 4 into a controlled extraction workflow. Language-model outputs are treated as provisional interpretations rather than independent literary judgements. Textual evidence, ontological constraints, and targeted expert adjudication govern how candidate relations enter the corpus-level representation.

5.1 Textual Anchoring and Bounded Neural Inference

The workflow begins by preserving two views of each verse: an *archival view* retaining the original text and a *normalised view* supporting consistent orthographic matching (Obeid et al., 2020). The verse or hemistich provides the principal extraction boundary, while poet, period, and genre metadata retain contextual information. Morphological analysis identifies features such as clitics, gender, number, and person, providing evidence for interpreting elliptical constructions and omitted participants (Pasha et al., 2014). These features inform, but do not determine, literary-semantic interpretation.

The language model receives an *ontology signature* specifying admissible participant classes, semantic relations, and controlled categories. It proposes event-centred relations grounded in textual spans, including inferred participants and figurative references. Such proposals remain provisional: conformity with the specified vocabulary does not establish that an interpretation is uniquely supported by the verse.

5.2 Deterministic Validation and Expert Adjudication

Candidate relations undergo deterministic checks for structural admissibility, including domain and range compatibility, required textual evidence, cardinality, and predefined semantic dependencies. SHACL and associated logical rules identify violations that may require rejection, deterministic repair, or human review. These checks establish conformity with the formal representation, not literary correctness.

Low-confidence assertions, unresolved contradictions, and explicitly flagged interpretations are referred for Human-in-the-Loop (HITL) adjudication. Reviewers examine the proposed relations alongside their textual evidence and may confirm, modify, reject, or retain them as unresolved. Presenting machine-generated relations as provisional hypotheses is intended to limit undue reliance on automated suggestions (Schroeder et al., 2025). Where competing readings remain defensible, uncertainty and reviewer dispositions are preserved rather than replaced by artificial agreement.

Accepted or adjudicated assertions are incorporated into an evidence-linked knowledge graph retaining their provenance, confidence status, and validation outcomes (Stranisci et al., 2023). This supports inspection of the textual basis for subsequent corpus-level findings. Technical specifications, validation procedures, and execution parameters are documented in Supplementary Note S3, Supplementary Algorithm S1, and Supplementary Fig. S3.

5.3 Illustrative Case Study: Figurative Ambiguity in Verse

Figure 1 illustrates the workflow through a verse attributed to the eleventh-century Andalusian poet Ibn Hamdis (Record ID: 3446). The example demonstrates the representation procedure rather than constituting an independent evaluation result.

The verse depicts a departure scene in which the female beloved sighs, turns away, and departs, while the male speaker experiences an emotional consequence. Contextual analysis identifies the sequential actions and resolves the elliptical participant structure, representing the female figure as an Agent and the speaker as an Experiencer. Within the study's operational categories, the resulting relations include Active agency, Physical objectification, and Male-to-Female gaze.

The second hemistich introduces metaphorical imagery involving the speaker's heart. Rather than treating this imagery as an unambiguous literal spatial event, the representation retains medium confidence and a review flag. The resulting graph links the proposed interpretation to its textual evidence while preserving the need for contextual scholarly judgement. The example illustrates how structured extraction can make a literary reading inspectable without presenting formal validation as a resolution of poetic ambiguity.

6 Empirical Validation and Corpus-Scale Literary Analysis

The evaluation addresses two connected questions: whether MARAYA produces sufficiently reliable, evidence-linked representations of literary relations, and what those representations reveal about gendered agency and objectification in the analysed poetry. The first question requires comparison with expert annotations and assessment of formal consistency. The second requires interpretation of corpus-level distributions within the conventions of Classical Arabic poetic discourse. Neither formal validity nor agreement with an annotation scheme is treated as proof of a uniquely correct literary reading.

6.1 Corpus Stratification, Annotation, and Inter-Annotator Agreement

The study analysed 31,209 single-verse poems from the Diwan collection (Al-Onazi et al., 2025), spanning twelve historical periods and eight genres. A proportional stratified sample of 4,500 verses was selected for expert annotation, with individual-poet contributions limited to reduce author-specific concentration. The annotated material comprised 18,742 event instances and 41,386 semantic assertions.

Four domain specialists followed an annotation protocol separating textual evidence spans from ontology-grounded semantic relations. After training and a 300-verse calibration exercise, 25% of the annotated corpus was independently double-annotated. Disagreements were logged and resolved through expert consensus; 12.7% of double-annotated instances required formal adjudication. This rate indicates that some cases required further

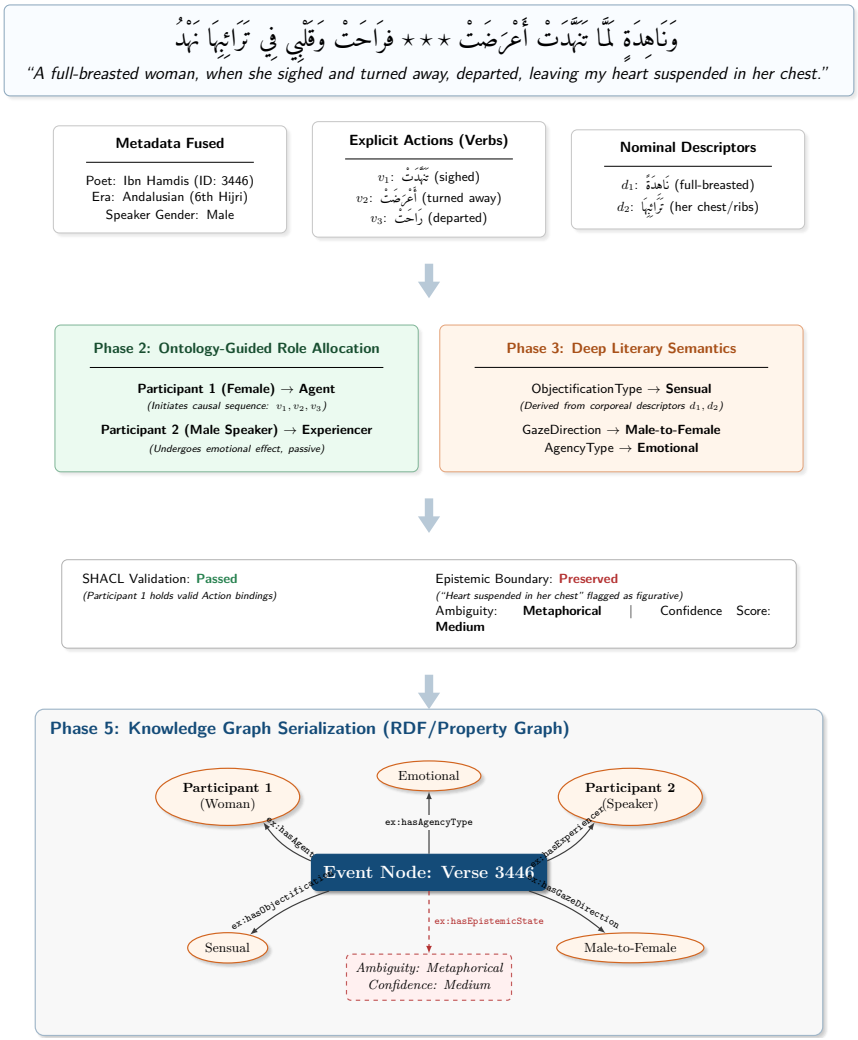


Figure 1: Semantic representation of Verse 3446 (Ibn Hamdis), showing textual anchoring, contextual inference, constraint validation, and the retention of metaphorical uncertainty for expert review.

interpretive review but does not by itself distinguish literary ambiguity from annotation disagreement. Overall agreement was Cohen's $\kappa = 0.85$ and Krippendorff's $\alpha = 0.84$ (Table 2). Agreement was higher for gender representation than for objectification. These metrics support consistent application of the operational schema but do not eliminate literary ambiguity.

The development data supported prompt calibration and constraint refinement. A separate, poet-disjoint 900-verse hold-out set was reserved for final evaluation. Execution settings are reported in Supplementary Table S1.

6.2 Constraint Enforcement and Extraction Reliability

Formal validation assessed 38,462 candidate assertions against 27 SHACL constraints and 11 ontology-specific logical rules. The Constraint Satisfaction Rate (CSR) measures the proportion of applicable constraints satisfied; the Contradiction Rate (CR) measures the proportion of evaluated assertions identified as contradictory. Validation increased CSR from 91.8% to 99.3% and reduced CR from 6.7% to 0.5% (Table 3). These changes indicate improved structural admissibility, not independently established interpretive correctness.

Expert-reference evaluation distinguished event detection from semantic classification. Automated semantic extraction achieved an F1 of 0.884 on the hold-out set; event detection achieved 0.912. Following targeted expert adjudication, the curated representation reached an F1 of 0.898. This latter value describes the agreement of the expert-assisted resource with the reference, not an improvement in the predictive accuracy of the unaided model.

Table 2: Corpus characteristics and inter-annotator agreement.

Corpus statistics		**Inter-annotator agreement**		
Characteristic	Value	Dimension	κ	α
Verses analysed	31,209	Event identification	0.89	0.88
Annotated verses	4,500	Semantic roles	0.86	0.85
Event instances	18,742	Agency status	0.83	0.82
Semantic assertions	41,386	Gender representation	0.91	0.90
Periods / genres	12 / 8	Objectification type	0.79	0.78
Events per verse	4.16	**Overall**	**0.85**	**0.84**

Table 3: Logical consistency and violation profile before and after validation.

Metric	**Before**	**After**	**Primary Violation Category**	**Count**	**Share**
Constraint Satisfaction Rate	91.8%	99.3%	Role Compatibility (e.g., Agency conflict)	1,248	39.6%
Contradiction Rate	6.7%	0.5%	Missing Evidence Spans	675	21.4%
Valid Assertions	35,310	37,605	Cardinality Violations	536	17.0%
Flagged for Review	2,181	661	Ontology Class Conflicts	406	12.9%
Rejected Assertions	971	196	Invalid Relation Types	287	9.1%

Category-level performance was highest for gender representation (F1=0.941) and lowest for objectification (F1=0.830). The difference matters substantively: identifying a gender-coded participant is often more directly anchored in linguistic evidence than deciding whether a metaphor, relational description, or evaluative expression constitutes objectification under the study's definitions.

Table 4: Expert-reference validation: overall results and semantic-category performance.

Evaluation task or category	Precision	Recall	F1	κ
Overall evaluation				
Event detection	0.921	0.904	0.912	0.883
Automated semantic extraction	0.893	0.876	0.884	0.846
Post-HITL curated representation	0.907	0.890	0.898	0.864
Automated semantic-category evaluation				
Gender representation	0.948	0.934	0.941	–
Agency status	0.887	0.861	0.874	–
Objectification type	0.842	0.819	0.830	–
Gaze relations	0.876	0.853	0.864	–

The ablation comparison further distinguishes the contribution of successive processing layers (Table 5). Adding ontology guidance increased F1 from 0.767 to 0.833; introducing deterministic constraints raised it to 0.884. The corresponding gains were 6.6 and 5.1 percentage points. The additional 1.4-point difference after expert adjudication reflects curation rather than a fourth automated model. Granular agency metrics by class (Supplementary Table S2), marginal gains, repeated-run stability, and graph-scale statistics are documented in Supplementary Note S4.

6.3 Corpus-Scale Patterns of Gendered Agency and Objectification

The resulting knowledge graph contains 208,341 nodes and 470,354 edges (see Supplementary Table S3 for comprehensive topology), providing a structured basis for examining recurring participant relations across the corpus. Its value for literary inquiry lies less in graph size than in the ability to trace aggregate patterns back to particular verses and their recorded interpretations.

The clearest distributional difference concerns agency. Active agency accounts for 61.8% of assignments within the male-coded group, compared with 48.2% within the female-coded group: a difference of 13.6 percentage points. Female-coded participants have correspondingly higher reactive (23.7% versus 18.5%) and passive (19.4% versus 12.1%) shares (Table 7).

Table 5: Ablation results across automated configurations and the expert-curated reference.

Configuration	P	R	F1	κ	CSR	CR
A: LLM only	.781	.754	.767	.692	83.4%	12.8%
B: Ontology-guided	.846	.821	.833	.781	91.6%	6.9%
C: Ontology + constraints	.893	.876	.884	.846	99.3%	0.5%
D: Post-HITL corpus	.907	.890	.898	.864	99.6%	0.2%

These differences can be read alongside the organisation of voice and action in Classical Arabic poetry. The conventions of the *qasida*, particularly its journeying and self-assertive passages, and those of *fakhr*, offer established positions for a speaking subject who initiates movement, declares intention, and performs or recounts action. The *ghazal*, meanwhile, frequently organises desire around relations of approach, separation, remembrance, and response. Within such configurations, a beloved may exercise agency through departure or withholding (*hajr*) while simultaneously functioning as the focus of another speaker's narration.

The observed distributions are compatible with an unequal allocation of initiating and responsive positions within these literary conventions. They do not, however, demonstrate that every male-coded participant is a speaking subject or that every female-coded participant is a passive beloved. Nor do they establish that the genres themselves cause the measured disparity. Such claims would require genre-specific comparisons and close readings beyond the aggregate results reported here. The finding is therefore a patterned asymmetry in represented agency, not a direct measurement of the social autonomy of historical men and women.

Table 6: Gender-based network characteristics of represented participants.

Metric	Male-coded	Female-coded
Degree centrality	0.238	0.191
Betweenness centrality	0.124	0.089
Clustering coefficient	0.318	0.354
Average interaction count	5.82	4.37

The network measures offer a complementary description. Male-coded participants exhibit higher reported degree and betweenness centrality, whereas female-coded participants exhibit a higher clustering coefficient (Table 6). These measures suggest differences in how participants are connected within the constructed semantic network. They should not be equated directly with narrative importance, psychological influence, or social power: each metric depends on the graph's representational choices and the relations included.

Table 7: Corpus-scale distributions of agency and objectification.

Gender-coded group	Active	Reactive	Passive	Ambiguous
Male-coded	61.8%	18.5%	12.1%	7.6%
Female-coded	48.2%	23.7%	19.4%	8.7%
Objectification modes				
Symbolic		38.4%		
Emotional		27.1%		
Physical		21.6%		
Mixed		12.9%		

Note. Agency percentages use the total classified agency assignments within each resolved gender-coded group as their respective denominators; unknown-gender assignments are excluded. Objectification percentages use all classified objectification events as their common denominator. The two distributions therefore describe different analytical populations.

A notable distributional difference appears in the modes of objectification. Symbolic objectification accounts for 38.4% of classified instances, compared with 21.6% for physical objectification. This 16.8-percentage-point difference indicates that symbolic objectification was more frequent under the study's classification scheme.

Classical Arabic verse employs devices such as *tashbīh* (simile), *majāz* (figurative substitution), and *kināyah*

(metonymic indirection), alongside motifs such as the beloved as a gazelle or celestial body, the beloved's phantom (*ṭayf*), and ruined encampments (*aṭlāl*). The observed distribution alone cannot establish that these conventions account for the pattern. The findings concern *literary representation*, not direct evidence of the lived historical or social realities of classical Arab men and women.

The interpretive issue is not simply that figurative language is present. It is whether, in particular passages, the substitution makes a person primarily available as an image, emblem, or object of another speaker's desire while reducing the representation of that person's independent intentions. Symbolic objectification, as operationalised here, offers a way to investigate that question across many verses. Yet metaphor is not inherently objectifying: an image may also express admiration, loss, emotional reciprocity, or the beloved's power over the speaker. Consequently, the aggregate percentage cannot establish the function of any individual trope without examining its textual context.

The contrast between symbolic and physical categories therefore supports a bounded literary proposition: the corpus contains more objectifying events classified through symbolic mediation than through physical description. It does not establish that symbolic objectification is necessarily less restrictive, more aesthetically elevated, or culturally unique to Arabic poetry.

The historical metadata permits these relations to be viewed across the twelve represented periods (Fig. 2). The figure offers a starting point for identifying period-specific differences and recurring patterns. Without controlling for genre composition, poet representation, and the number of observations within each period, however, the displayed trajectories cannot independently establish continuity, change, or the historical causes of either.

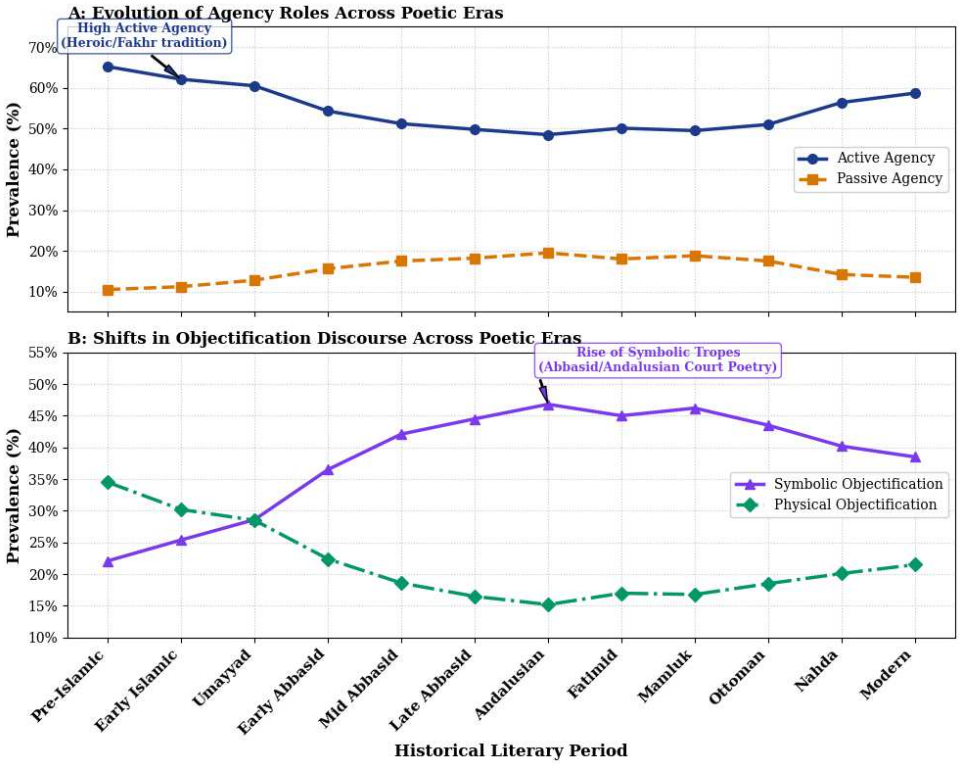


Figure 2: Temporal distribution of selected agency- and objectification-related categories across the historical periods represented in the corpus. The plotted distributions are descriptive and do not independently establish historical causation.

6.4 Epistemic Boundaries and Methodological Threats

Three limitations qualify these findings. First, agency, objectification, and gaze are interpretive constructs operationalised through a finite set of categories. Expert agreement and formal validation support consistent application but cannot remove the plurality of defensible readings. Second, automated extraction remains sensitive to contextual inference, especially metaphorical attribution and indirect agency; the lower performance for objectification and

ambiguous agency reflects this difficulty.

Third, the single-verse analytical boundary restricts access to relationships developed across an entire *qasida*. A departure, remembered encounter, or apparently passive response may acquire a different function within an extended sequence of verses. The present results therefore characterise locally represented relations, not complete poetic narratives. The corpus also cannot establish equivalent distributions for unrepresented texts, genres, or traditions.

Within these boundaries, MARAYA provides an inspectable basis for corpus-scale literary comparison. Its contribution is to make recurrent patterns of gendered representation available for critical examination while keeping statistical description, formal admissibility, and literary interpretation distinct.

7 Error Analysis and Interpretive Boundaries

The interpretive value of computational literary analysis depends not only on the reliability of its classifications but also on recognising what those classifications cannot establish. Classical Arabic poetry presents particular difficulties through compressed expression, layered metaphor, omitted arguments, and historically situated allusions (Alsiyat and Piao, 2020; Aloraini et al., 2022; Al Ghallabi et al., 2025). These difficulties are especially consequential when computationally inferred participants and actions become the basis for claims about gender, agency, objectification, and gaze. Examining residual errors therefore provides a means of identifying where structured representation encounters the limits of local textual evidence and formal reasoning.

7.1 Hermeneutic Sources of Extraction Error

We analysed 400 randomly sampled false-positive and false-negative extraction errors from the poet-disjoint hold-out set of 900 verses, comprising approximately 8,280 semantic assertions. Each error was assigned a primary cause, although the underlying interpretive difficulties may overlap. Table 8 reports the complete taxonomy.

Table 8: Taxonomy and distribution of analysed extraction errors ($N = 400$).

Error category and specific failure mode	Count	Category share	Total share
1. Metaphorical Density	**168**	**100.0%**	**42.0%**
Event Hallucination (literalising figurative imagery)	104	61.9%	26.0%
Agency Misclassification (rhetorical subject agency)	64	38.1%	16.0%
2. Coreference and Elliptical Ambiguity	**112**	**100.0%**	**28.0%**
Implicit and Elliptical Reference (zero-anaphora)	72	64.3%	18.0%
Gender Resolution Failures (conflicting agreement cues)	40	35.7%	10.0%
3. Cultural and Historical Context Dependence	**84**	**100.0%**	**21.0%**
Historical Allusions	38	45.2%	9.5%
Genealogical and Entity References	26	31.0%	6.5%
Cultural Symbolism and Conventional Motifs	20	23.8%	5.0%
4. Minor Syntactic and Other Failures	**36**	**100.0%**	**9.0%**
Total Analysed Errors	**400**	–	**100.0%**

Metaphorical Density accounted for 168 errors (42.0%), the largest category. In 104 cases, figurative imagery was represented as a concrete event; in another 64, agency was assigned incorrectly to a rhetorical subject. These failures expose a difficulty specific to literary-semantic extraction: grammatical participation in an expression does not necessarily imply agency within its figurative meaning. Literalising personification may create an event that the

poem does not assert, while assigning agency from surface syntax may obscure the relation between a speaker, an image, and the participant being represented. In 75 metaphor-related cases (44.6%), the initial error affected at least one additional semantic component. Such propagation matters because a mistaken event interpretation can alter subsequent claims about agency or objectification.

Coreference and Elliptical Ambiguity accounted for 112 errors (28.0%), including 72 involving implicit or elliptical reference and 40 involving gender resolution. Arabic poetic compression permits subjects and participants to remain unexpressed, while changes in address or speaker perspective may complicate their identification. A mistaken antecedent can consequently affect several relations constructed around the same participant. This occurred in 65 cases (58.0%), where reference-resolution errors propagated to another semantic component. The result illustrates why locally plausible grammatical decisions may remain insufficient for interpreting agency and gaze within a broader poetic utterance.

Cultural and Historical Context Dependence accounted for 84 errors (21.0%). Historical allusions, genealogical references, and conventional motifs may require knowledge unavailable within the isolated verse. In these cases, the problem is not necessarily malformed extraction but an interpretation that remains incomplete or misleading without its literary and historical setting. The remaining 36 errors (9.0%) involved minor syntactic and other failures. Supplementary Fig. S4 examines the propagation mechanisms associated with these categories.

7.2 Formal Consistency versus Interpretive Validity

The clearest evidence of the distinction between computational admissibility and literary interpretation emerged from the context-dependent errors. Of the 84 errors attributed to cultural and historical context, 53 (63.1%) satisfied the applicable SHACL shapes, cardinality requirements, and ontology constraints despite diverging from expert-adjudicated interpretations.

This finding identifies a substantive epistemic boundary. Formal validation determines whether an assertion conforms to explicitly encoded representational conditions. It can reject incompatible participant roles, missing required relations, and structural contradictions. It cannot, without additional interpretive evidence, establish whether an allusion has been recognised, whether a conventional image has been read appropriately, or whether a historically situated relation has been understood in its literary context.

The distinction is particularly important for the present study because the resulting representations support aggregate claims about gendered agency and objectification. A formally valid but contextually mistaken assertion can enter a statistical distribution unless its interpretive uncertainty is recognised. High constraint satisfaction must therefore not be confused with equally high hermeneutic validity.

Expert adjudication provides an essential corrective, but it does not turn interpretation into an infallible procedure. Scholarly readings may themselves be contested, especially where poetic ambiguity is deliberate. The appropriate objective is consequently not to eliminate interpretive plurality but to distinguish supported classifications, unresolved alternatives, and representations requiring revision. MARAYA's evidentiary contribution lies in making these decisions traceable rather than presenting computational output as an exhaustive account of literary meaning.

7.3 Scope, Limitations, and Bounded Generalisability

The findings remain conditional on the evaluated corpus, its annotation scheme, and the extraction boundaries adopted. The single-verse unit provides a manageable context for textual anchoring and limits the complexity of cross-verse reference resolution. It also excludes information distributed across the complete *qasida*, including changes in speaker perspective, narrative progression, and the development of relationships between participants. Accordingly, the reported agency and objectification patterns describe locally represented relations rather than complete poetic narratives.

The reliance on an externally provided commercial language model introduces a further reproducibility limitation. Model implementations may change independently of the documented prompts and validation rules, potentially affecting candidate interpretations. Moreover, the ontology's categories and adjudication procedures

reflect particular scholarly decisions. Strong agreement within that framework does not establish that alternative critical traditions would organise every passage in the same way.

Finally, the framework was evaluated on Classical Arabic poetry, not on modern free verse, historical prose, dialectal literature, or non-Arabic corpora. Transfer to these domains would require renewed annotation, adaptation of representational categories, and independent evaluation. Supplementary Table S4 distinguishes demonstrated applicability from prospective extensions.

Future work should therefore prioritise discourse-level participant tracking, historically grounded reference resolution, refinement of ambiguous categories, and interfaces that expose competing interpretations to expert reviewers. These directions are detailed in Supplementary Note S5. Within its present boundaries, MARAYA offers an inspectable instrument for corpus-scale literary inquiry, not a substitute for contextual reading or scholarly judgement.

8 Conclusion

This study examined how gendered agency and objectification are represented across a large historical corpus of Classical Arabic poetry. Analysis of 31,209 single-verse poems revealed corpus-level differences in represented agency, with male-coded participants showing a higher proportion of active agency and female-coded participants showing higher proportions of reactive and passive involvement, while symbolic objectification was more frequent than direct physical reduction (RQ3). These findings provide corpus-level evidence of patterned gender representation within the analysed corpus. MARAYA provides the semantic and evidential infrastructure through which such relations can be represented in an inspectable form (RQ1), while formal ontological constraints improve the structural validity of the extracted representations (RQ2).

The humanistic contribution of the study lies in providing corpus-level empirical evidence of patterned gendered agency and objectification within the analysed Classical Arabic poetry, while retaining their connection to individual textual evidence. The computational framework serves as the methodological means for making these patterns inspectable at corpus scale; computational validation supports the reliability and structural consistency of the extraction process, but does not substitute for historically situated literary interpretation or establish a single authoritative reading of the texts. The single-verse structure likewise limits the analysis to localised semantic relations and does not capture broader narrative or thematic development across longer poems. Accordingly, MARAYA should be understood as an evidence-grounded instrument for extending the scale and transparency of literary analysis, with its applicability to other figurative or historical corpora remaining prospective and requiring domain-specific revalidation.

Declarations

Acknowledgements This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2026R263), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Ethics Declaration The study used publicly available historical and literary texts as its corpus. Four domain specialists participated in annotation, independent coding, and expert adjudication; no clinical, sensitive personal, or identifiable participant data were collected. The authors confirm that the annotation and expert-validation procedures complied with the applicable institutional requirements for this type of research.

Conflict of Interest The authors declare no competing interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data and Code Availability The dataset, annotated corpus, and complete extraction framework (source code) generated and analysed during this study are publicly available in the Zenodo repository: https://doi.org/10.5281/zenodo.21098754.

Author Contributions Conceptualization, B.B.A., H.A., A.A.A., and W.A.N.; methodology, W.A.N., B.B.A., and A.A.A.; software, W.A.N.; validation, A.M.A., W.F.B., N.I.A., and R.G.A.; formal analysis, W.A.N. and A.A.A.; investigation, R.G.A., N.I.A., W.F.B., and A.M.A.; resources, B.B.A. and H.A.; data curation, W.A.N., R.G.A., and W.F.B.; writing—original draft preparation, A.A.A., W.A.N., and B.B.A.; writing—review and editing, H.A., B.B.A., A.A.A., W.A.N., A.M.A., W.F.B., N.I.A., and R.G.A.; visualization, W.A.N.; supervision, B.B.A. and H.A.; project administration, B.B.A. and H.A.; funding acquisition, B.B.A. and H.A. All authors have read and agreed to the published version of the manuscript.

Language editing assistance AI-assisted language tools were used solely for grammatical proofreading and vocabulary polishing of the authors' original text, as English is not their first language. No generative text creation or conceptual writing was performed by AI. All ideas, analyses, and written arguments were conceived and authored entirely by the researchers, who take full responsibility for the final text.

References

Haslhofer, B., Isaac, A., and Simon, R. (2018). Knowledge graphs in the libraries and digital humanities domain. *arXiv preprint arXiv:1803.03198*. https://doi.org/10.1007/978-3-319-63962-8_291-1

Huang, Y. Y. et al. (2023). Using knowledge graphs and deep learning algorithms to enhance digital cultural heritage management. *Heritage Science*, 11, Art. 204. https://doi.org/10.1186/s40494-023-01042-y

Schilling-Wilhelmi, M. et al. (2025). From text to insight: Large language models for chemical data extraction. *Chemical Society Reviews*, 54, 1125–1150. https://doi.org/10.1039/D4CS00913D

Tong, X., Choenni, R., Lewis, M., and Shutova, E. (2024). Metaphor understanding challenge dataset for LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3517–3536. https://doi.org/10.18653/v1/2024.acl-long.193

Wachowiak, L. and Gromann, D. (2023). Does GPT-3 grasp metaphors? Identifying metaphor mappings with generative language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1018–1032. https://doi.org/10.18653/v1/2023.acl-long.58

Huang, L. et al. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 1–55. https://doi.org/10.1145/3703155

Zaman, G. et al. (2021). An ontological framework for information extraction from diverse scientific sources. *IEEE Access*, 9, 42111–42124. https://doi.org/10.1109/ACCESS.2021.3063181

Geng, S., Josifoski, M., Peyrard, M., and West, R. (2023). Grammar-constrained decoding for structured NLP tasks without finetuning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10932–10952. https://doi.org/10.18653/v1/2023.emnlp-main.674

Darwish, K. et al. (2021). A panoramic survey of natural language processing in the Arab world. *Communications of the ACM*, 64(4), 72–81. https://doi.org/10.1145/3447735

Farghaly, A. and Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing*, 8(4), Art. 14, 1–22. https://doi.org/10.1145/1644879.1644881

Aloraini, A., Yu, J., Aliady, W. A., and Poesio, M. (2025). A survey of coreference and zeros resolution for Arabic. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 24(10), 1–24. https://doi.org/10.1145/3702323

Abd Alhadi, H., Hussein, A. A., and Kuflik, T. (2023). Automatic identification of rhetorical elements in classical Arabic poetry. *Digital Humanities Quarterly*, 17(3). https://doi.org/10.63744/tax65tcjkn2p

Abo, M. E. M., Raj, R. G., and Qazi, A. (2019). A review on Arabic sentiment analysis: State-of-the-art, taxonomy and open research challenges. *IEEE Access*, 7, 162008–162024. https://doi.org/10.1109/ACCESS.2019.2951530

El Moussaoui, T. and Loqman, C. (2024). Advancements in Arabic named entity recognition: A comprehensive review. *IEEE Access*, 12, 180238–180266. https://doi.org/10.1109/ACCESS.2024.3491897

Al-Onazi, B. B. et al. (2025). Diwan": Constructing the largest annotated corpus for Arabic poetry. *IEEE Access*, 13, 58927–58941. https://doi.org/10.1109/ACCESS.2025.3551161

Al-Anazi, R. G., Nashir, W. A., Al-Onazi, B. B., Alhamad, A. A., and Al-Shargabi, A. A. (2025). Maqasid: A hybrid CNN-BiLSTM framework for nuanced thematic classification of Arabic poetry. *IEEE Access*, 13, 189421–189443. https://doi.org/10.1109/ACCESS.2025.3621112

Shoubaki, H., Abdallah, S., and Shaalan, K. (2024). Deep learning techniques for identifying poets in Arabic poetry: A focus on LSTM and Bi-LSTM. *Procedia Computer Science*, 244, 461–470. https://doi.org/10.1016/j.procs.2024.10.221

Alsmearat, K., Al-Ayyoub, M., Al-Shalabi, R., and Kanaan, G. (2017). Author gender identification from Arabic text. *Journal of Information Security and Applications*, 35, 85–95. https://doi.org/10.1016/j.jisa.2017.06.003

Karimullah, K. (2020). Sketching women: A corpus-based approach to representations of women's agency in political Internet corpora in Arabic and English. *Corpora*, 15(1), 21–53. https://doi.org/10.3366/cor.2020.0184

Stranisci, M. A. et al. (2023). The World Literature Knowledge Graph. In *The Semantic Web – ISWC 2023*, pages 435–452. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47243-5_24

Meghini, C., Bartalesi, V., and Metilli, D. (2021). Representing narratives in digital libraries: The narrative ontology. *Semantic Web*, 12(2), 241–264. https://doi.org/10.3233/SW-200421

Arsenyan, V. et al. (2024). Large language models for biomedical knowledge graph construction: Information extraction from EMR notes. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 295–317. https://doi.org/10.18653/v1/2024.bionlp-1.23

Kuculo, T., Abdollahi, S., and Gottschalk, S. (2025). Transformer-based architectures versus large language models in semantic event extraction: Evaluating strengths and limitations. *Semantic Web*, 16(5). https://doi.org/10.1177/22104968251363759

Al-Anazi, R. G. et al. (2026a). Deep Rhetoric: A Qarīna-Aware Multi-Task Learning Framework for Hierarchical Analysis of Qur'anic Metaphor. *IEEE Access*, 14. https://doi.org/10.1109/ACCESS.2026.3696593

Zibin, A., Binhaidara, N., Al-Shahwan, H., and Yousef, H. (2025). Metaphor interpretation in Jordanian Arabic, Emirati Arabic and Classical Arabic: Artificial intelligence vs. humans. *Humanities and Social Sciences Communications*, 12(1), 1–12. https://doi.org/10.1057/s41599-025-05282-0

Habash, N. et al. (2022). Camel Treebank: An open multi-genre Arabic dependency treebank. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2672–2681. https://doi.org/10.63317/23t54mrdwiiy

Nashir, W. A. et al. (2025a). A complete, multi-layered quranic treebank dataset with hybrid syntactic annotations for classical arabic processing. *Data in Brief*, 62, 111940. https://doi.org/10.1016/j.dib.2025.111940

Sims, M., Park, J. H., and Bamman, D. (2019). Literary event detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3623–3634. https://doi.org/10.18653/v1/P19-1353

Nashir, W. A. et al. (2025b). Noor: A Pipeline Framework for Classical Arabic Parsing. *IEEE Access*, 13, 167538–167559. https://doi.org/10.1109/ACCESS.2025.3607915

Mutawa, A. M. and Alrumaih, A. (2025). Determining the meter of classical Arabic poetry using deep learning: A performance analysis. *Frontiers in Artificial Intelligence*, 8, Art. 1523336. https://doi.org/10.3389/frai.2025.1523336

Nashir, W. A. et al. (2026b). Multi-Label Classification of Qur'anic Similes: A Computational Approach to Arabic Rhetorical Theory. *Sana'a University Journal of Applied Sciences and Technology (JAST)*, 4(5), 2064–2074. https://doi.org/10.59628/jast.v4i5.2677

Al-Anazi, R. G. et al. (2026b). Burhan: A Constraint-Based Engineering Framework for Computational Rhetorical Semantics in Classical Arabic. *IEEE Access*, 14. https://doi.org/10.1109/ACCESS.2026.3696591

Aladeemy, A. A. et al. (2024). Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development. *Heliyon*, 10(21), Art. e39786. https://doi.org/10.1016/j.heliyon.2024.e39786

Qu, X. et al. (2024). A survey on Arabic named entity recognition: Past, recent advances, and future trends. *IEEE Transactions on Knowledge and Data Engineering*, 36(3), 943–959. https://doi.org/10.1109/TKDE.2023.3303136

Qarah, F. (2024). AraPoemBERT: A pretrained language model for Arabic poetry analysis. *arXiv preprint arXiv:2403.12392*.

Kalathil, S. S., Li, T., Dong, H., and Liang, H. (2024). HTEKG: A human-trait-enhanced literary knowledge graph with language model evaluation. In *Proc. 16th Int. Joint Conf. Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024)*, volume 2, pages 207–215. https://doi.org/10.5220/0013013600003838

Shiri, F. et al. (2024). Decompose, enrich, and extract! Schema-aware event extraction using LLMs. In *Proc. 27th Int. Conf. Information Fusion (FUSION)*, pages 1–8. https://doi.org/10.23919/FUSION59988.2024.10706385

Zhou, C. et al. (2026). TextMineX: Data, evaluation framework and ontology-guided LLM pipeline for humanitarian mine action. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 2505–2523. https://doi.org/10.18653/v1/2026.findings-eacl.130

Shaalan, K. (2014). A survey of Arabic named entity recognition and classification. *Computational Linguistics*, 40(2), 469–510. https://doi.org/10.1162/COLI_a_00178

Kuhn, J. (2019). Computational text analysis within the Humanities: How to combine working practices from the contributing fields? *Language Resources and Evaluation*, 53(4), 565–602. https://doi.org/10.1007/s10579-019-09459-3

Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43(5–6), 625–640. https://doi.org/10.1006/ijhc.1995.1066

Ciotti, F. (2016). Toward a formal ontology for narrative. *MATLIT: Materialities of Literature*, 4(1), 29–44. https://doi.org/10.14195/2182-8830_4-1_2

Chen, X. et al. (2019). Embedding uncertain knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3363–3370. https://doi.org/10.1609/aaai.v33i01.33013363

Ferranti, N., De Souza, J. F., Ahmetaj, S., and Polleres, A. (2024). Formalizing and validating Wikidata's property constraints using SHACL and SPARQL. *Semantic Web*, 15(6), 2333–2380. https://doi.org/10.3233/SW-243611

Brown, S. W., Bonial, C., Obrst, L., and Palmer, M. (2017). The rich event ontology. In *Proceedings of the Events and Stories in the News Workshop*, pages 87–97. https://doi.org/10.18653/v1/W17-2712

Ehrmann, M. et al. (2023). Named entity recognition and classification in historical documents: A survey. *ACM Computing Surveys*, 56(2), 1–47. https://doi.org/10.1145/3604931

Stuhler, O. (2024). The gender agency gap in fiction writing (1850 to 2010). *Proceedings of the National Academy of Sciences*, 121(30). https://doi.org/10.1073/pnas.2319514121

Luo, K., Mao, Y., Zhang, B., and Hao, S. (2024). Reflecting the male gaze: Quantifying female objectification in 19th and 20th century novels. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13803–13812. https://doi.org/10.63317/5d5td7gmuc2b

Pareti, P. and Konstantinidis, G. (2022). A review of SHACL: From data validation to schema reasoning for RDF graphs. In M. Šimkus and I. Varzinczak, editors, *Reasoning Web. Declarative Artificial Intelligence*, pages 88–120. Springer. https://doi.org/10.1007/978-3-030-95481-9_6

Obeid, O. et al. (2020). CAMeL Tools: An open source Python toolkit for Arabic natural language processing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 7022–7032.

Pasha, A. et al. (2014). MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1094–1101. https://doi.org/10.63317/3piegahzyagx

Schroeder, H., Roy, D., and Kabbara, J. (2025). Just put a human in the loop? Investigating LLM-assisted annotation for subjective tasks. In *Findings of the Association for Computational Linguistics: ACL 2025*. https://doi.org/10.18653/v1/2025.findings-acl.1323

Alsiyat, I. and Piao, S. S. L. (2020). Metaphorical expressions in automatic Arabic sentiment analysis. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4911–4916.

Aloraini, A., Pradhan, S., and Poesio, M. (2022). Joint coreference resolution for zeros and non-zeros in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 177–187. https://doi.org/10.18653/v1/2022.wanlp-1.2

Chen, H., Wang, J., and Zhu, L. (2024). The representation of the body in Charles Dickens's novels: a computational approach. *Humanities and Social Sciences Communications*, 11(1), Art. 233. https://doi.org/10.1057/s41599-024-03366-x

Al Ghallabi, W. et al. (2025). Fann or Flop: A multigenre, multiera benchmark for Arabic poetry understanding in LLMs. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. https://doi.org/10.18653/v1/2025.emnlp-main.1023

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf