Comparing LLM-Based Knowledge Graph Extraction Approaches on Literary Studies in Spanish: A Case Study on Orbis Tertius

Federico Gabriel Cortes
Bergische Universität Wuppertal
Gaußstraße 20, Wuppertal
fede.gcortes@gmail.com

Abstract

Knowledge graph construction from scholarly text increasingly relies on large language models, yet different extraction architectures produce different graphs. Literary studies poses particular challenges: meaning is interpretive rather than factual, and the boundaries of relevant knowledge are determined by hermeneutic frameworks rather than empirical verification. We compare two LLM-based extraction frameworks—entity-anchored extraction (KGGen) and open extraction with schema canonicalization (EDC)—on 472 Spanish-language literary studies articles from Orbis Tertius (1996–2024). Despite fundamental architectural differences, both methods converge on key findings: cultural framing dominates literary discourse by 2.2–2.5$\times$ over textual framing ($p < .001$), and core author networks remain consistent across approaches. The methods diverge in entity composition: KGGen captures more proper names (40.7% vs. 18.7%), while EDC captures more abstract concepts (42.8%) and preserves Spanish predicates with 21,025 semantic definitions. Convergent findings across architecturally different methods merit higher confidence, and we identify methodological considerations for knowledge graph construction from humanities scholarship.

Keywords: knowledge graph extraction, literary studies, LLM, Spanish NLP, digital humanities

1. Introduction

Knowledge graphs have become a fundamental data structure for organizing interconnected information, with applications ranging from question-answering to recommendation systems (Ji et al., 2022). Knowledge graph construction from scholarly text increasingly relies on generative methods (Ye et al., 2023; Bian, 2025), yet different extraction architectures produce different graphs. This raises a methodological question for researchers building knowledge graphs from academic literature: do findings depend on the extraction approach, or do they reflect genuine patterns in the corpus?

Traditional named entity recognition (NER) identifies persons, locations, and organizations—categories developed for news and general text. However, humanities scholarship includes theoretical concepts (*deconstruction*, *modernity*), aesthetic categories (*sublime*, *minor literature*), and disciplinary vocabulary (*writing*, *poetics*) that fall outside NER ontologies. Recent studies have demonstrated that LLMs exhibit strong performance on open information extraction tasks (Li et al., 2023), suggesting the potential to capture vocabulary that traditional NER misses. However, different LLM-based architectures may capture this vocabulary differently, and the resulting knowledge graphs may exhibit different structural properties.

Yet the challenge runs deeper than vocabulary. In empirical domains, extraction can be evaluated against verifiable facts: a protein-protein interaction either exists or does not, and a triplet can be judged correct or incorrect accordingly. In literary studies, language functions simultaneously as the object of study and the medium of analysis: a critic writing about Borges's *escritura* is not merely referencing a concept but performing a hermeneutic act in which the critic's own language is entangled with what it describes. Two articles may produce the same triplet—"Borges" linked to "gauchesca"—while making opposite interpretive claims. Knowledge graphs flatten this entanglement into subject-predicate-object structures, which is both their utility (they reveal aggregate discursive patterns) and their fundamental limitation (they lose interpretive force). This epistemological condition means that "correct" extraction is underdefined for literary scholarship, and no gold-standard evaluation is available. In such a setting, no single extraction method can claim correctness, but convergence across architecturally different methods provides a form of methodological triangulation (Heesen et al., 2019): if systems with different constraints and biases produce the same aggregate pattern, confidence increases that the pattern reflects the corpus rather than an artifact of either method—even though convergence cannot validate individual triplets.

We address this question by comparing two recent LLM-based extraction frameworks on the same corpus. KGGen (Mo et al., 2025) implements entity-anchored extraction that first identifies entities, then extracts relations constrained to those entities, followed by entity resolution to reduce graph sparsity. EDC (Zhang and Soh, 2024) implements open information extraction that extracts unconstrained triplets, defines their semantics, then canonicalizes to a schema. These approaches embody different philosophies: KGGen constrains relations to pre-extracted entities, en-

Proceedings of the 3rd Int. Workshop on Natural Scientific Language Processing (NSLP 2026) @ LREC 2026, pages 78–87
12 May 2026.

suring groundedness but potentially missing conceptual vocabulary; EDC allows any noun phrase as relation argument, capturing more concepts but with noisier boundaries. Both represent advances over earlier open information extraction methods like OpenIE (Angeli et al., 2015); as surveys of the field note, such methods generally produce non-canonicalized outputs, leading to redundant and ambiguous knowledge base entries (Kamp et al., 2023).

Our corpus comprises 472 articles from Orbis Tertius (ISSN 1851-7811), a peer-reviewed literary studies journal published by Universidad Nacional de La Plata, Argentina (1996–2024). While knowledge graphs have been constructed from literary texts (Stranisci et al., 2023; Kalathil et al., 2024) and from scholarly papers using few-shot LLM approaches (Lan et al., 2025), no direct comparison of extraction architectures on literary scholarship exists. This corpus presents challenges absent from typical NER benchmarks: scholarly prose references theoretical concepts alongside named entities, and "relevant" entities depend on interpretive framework rather than factual verification.

Our research questions are: (1) How do entity-anchored and open extraction approaches differ in what they capture from literary studies? (2) Do substantive findings—such as how literary scholars frame "literature" conceptually—hold across methods? (3) What methodological recommendations follow from comparing extraction approaches on humanities text?

This paper makes four contributions. First, we provide the first direct comparison of KGGen and EDC architectures on humanities scholarship. Second, we present evidence that key findings (cultural framing ratio, author networks) converge across methods. Third, we characterize what each approach captures: KGGen favors proper names, EDC favors conceptual vocabulary. Fourth, we address how to choose extraction methods for scholarly text. Our comparison suggests that convergent findings across architecturally different methods—like the cultural framing ratio we report below—merit higher confidence than findings from single methods.

2. Corpus: Orbis Tertius

2.1. The Journal and Its Institutional Context

Orbis Tertius (ISSN 1851-7811) is a peer-reviewed journal of literary theory and criticism published by the Centro de Estudios de Teoría y Crítica Literaria, a dual-dependency research institute of the Universidad Nacional de La Plata (UNLP) and CONICET (Argentina's National Scientific and Technical Research Council). Founded in 1996 during the post-dictatorship reconstruction of Argentine academic institutions, the journal has become a cornerstone of literary scholarship in South America. The journal's name references Jorge Luis Borges' story "Tlön, Uqbar, Orbis Tertius," anchoring its identity in the Río de la Plata literary tradition while signaling an interest in the invention of worlds through fiction and theory.

The journal's founding date is significant. The 1990s marked a decisive phase of institutionalization in Argentine literary studies, driven by the strengthening of humanities programs at public universities following the restoration of democracy in 1983 (Gerbaudo, 2024; Patiño, 1997). Orbis Tertius emerged from this institutional consolidation, representing the professionalization of a field that had previously operated through cultural magazines rather than peer-reviewed academic publication.

2.2. Open Access and Regional Significance

Orbis Tertius was among the first Latin American humanities journals to transition to digital Open Access, adopting the Open Journal Systems (OJS) platform in 2006 (Unzurrunzaga et al., 2015). The journal adheres to the Diamond Open Access model: it charges no fees to authors and no fees to readers. This is significant for Latin American researchers, who often lack funding for publication fees required by European or North American journals. In Argentina, the majority of scientific journals adhere to Open Access principles (Beigel and Salatino, 2015), and Law 26,899 on the Creation of Open Access Institutional Repositories (2013) formalized the requirement for publicly funded research to be openly accessible.

Metric	Value
Articles	472
Temporal span	1996–2024 (28 years)
Language	Spanish
Approximate words	2.4 million
Text chunks	4,728
Thematic scope	Latin American literature

Table 1: Corpus characteristics.

2.3. Significance for NLP Research

This corpus presents characteristics that make it valuable for knowledge graph extraction research beyond standard benchmarks. First, Spanish NLP resources lag significantly behind English, particularly for scholarly text. No KG extraction benchmarks exist for literary studies in Spanish, and most existing evaluation datasets derive from news, Wikipedia, or biomedical domains. The Orbis Tertius corpus provides an opportunity to evaluate extraction methods on a domain and language combination that remains underexplored.

Second, literary studies is rarely used for information extraction evaluation. Unlike empirical STEM domains, where relations can often be verified against databases (protein-protein interactions, chemical compounds), hermeneutic scholarship in the humanities includes theoretical concepts (*deconstrucción* 'deconstruction', *modernismo* 'modernism'), aesthetic categories (*lo sublime* 'the sublime', *literatura menor* 'minor literature'), and disciplinary vocabulary (*escritura* 'writing', *poética* 'poetics') alongside named entities. What counts as a "relevant" entity depends on interpretive framework rather than factual verification. This poses distinctive challenges for extraction systems designed around news or empirical scientific text. Third, the corpus is fully open access, enabling complete replication of our experiments.

2.4. Preprocessing

Articles were chunked into segments of approximately 5,000 characters, yielding 4,728 chunks. Both extraction methods processed identical chunks to ensure comparability. BERTopic clustering (Grootendorst, 2022) assigned articles to nine thematic clusters, preserved as provenance metadata for topic-specific analysis.

3. Methods

3.1. Traditional NER Approach

To establish a baseline, we applied spaCy's Spanish model (`es_core_news_lg`) to the corpus (Honnibal et al., 2020). This model extracts named entities in standard categories: PER (persons), LOC (locations), ORG (organizations), and MISC (miscellaneous). Traditional NER successfully identifies proper names—"Jorge Luis Borges," "Buenos Aires," "Universidad Nacional de La Plata"—but systematically misses the conceptual vocabulary central to literary scholarship. Terms like *escritura*, *modernismo*, or *poética* are not recognized as entities despite their centrality to scholarly discourse.

To quantify this limitation, we compared spaCy NER output with LLM-based extraction (KGGen) across all 472 articles. As shown in Table 2, the methods produce substantially different outputs: spaCy extracts nearly twice as many entity mentions (502 per article vs. 262), but entity overlap is low (Jaccard similarity 19.1%). Crucially, KGGen captures over 26,000 conceptual terms—*literatura*, *escritura*, *cultura*, *poesía*, *sociedad*—that fall entirely outside NER's ontology. This limitation motivates the use of LLM-based extraction methods that can capture both named entities and conceptual terms.

Metric	spaCy NER	KGGen
Total entity mentions	236,905	123,730
Unique entities	92,441	72,382
Per article (avg)	502	262
Conceptual terms	0	26,069

Entity overlap: Jaccard similarity 19.1%

Table 2: Comparison of spaCy NER and LLM-based extraction.

3.2. KGGen: Entity-Anchored Extraction

KGGen addresses a fundamental challenge in automatic knowledge graph extraction: the sparsity problem. As Mo et al. (2025) observe, existing extractors lack effective mechanisms for entity resolution and relation normalization, leading to graphs with nearly as many unique relation types as edges—sparse, disconnected representations that limit utility for downstream tasks. KGGen implements a multi-stage extraction pipeline using DSPy prompt programming (Khattab et al., 2023) with Gemini 2.0 Flash (Gemini Team, 2024). In the first phase, the model performs unconstrained entity extraction, identifying any noun phrase it considers salient: persons, places, works, concepts, and institutions. No entity type constraints are imposed at this stage. In the second phase, the model extracts relations, but these are constrained to use only entities from the first phase. This architectural choice forces grounding: every extracted relation must connect pre-identified entities rather than allowing arbitrary noun phrases as arguments.

KGGen includes an entity and edge resolution stage designed to identify nodes referring to the same underlying entities and consolidate edges with equivalent meanings (Mo et al., 2025). Inspired by crowdsourcing strategies for entity resolution (Wang et al., 2012), this iterative algorithm uses embedding-based clustering (S-BERT) combined with LLM-based de-duplication to merge synonymous entities. However, we found that this clustering did not effectively merge Spanish name variants in our corpus—entities like "Borges," "Jorge Luis Borges," and "J.L. Borges" were not being grouped. This challenge with name variants parallels documented limitations of embedding-based canonicalization (Vashishth et al., 2018). The failure likely reflects S-BERT's English-dominant training data, which produces embeddings less sensitive to Spanish naming conventions such as compound surnames and variable use of first names. We therefore did not apply KGGen's built-in entity resolution, instead using manual curation with fuzzy string matching for entity canonicalization (see Section 3.5). The entity-first approach nonetheless ensures that the resulting graph is anchored on recognizable entities, preventing hallucinated relations between noun phrases the model might invent during relation extraction. This design produces a higher proportion of proper names (40.7% of top entities), with predicates normalized to English forms and no semantic definitions generated. From 4,728 chunks, KGGen extracted 104,500 raw triplets.

3.3. EDC: Open Extraction with Schema Canonicalization

EDC (Extract-Define-Canonicalize) implements a three-phase pipeline (Zhang and Soh, 2024). EDC supports two operational modes: Target Alignment, where extracted triplets are canonicalized against a predefined schema, and Self Canonicalization, where the system dynamically constructs its own schema from the extracted triplets. Since no predefined knowledge graph schema exists for literary studies in Spanish, we use EDC in Self Canonicalization mode. We use Mistral (Jiang et al., 2023) as the base LLM, selected for its open-source availability and competitive performance on multilingual text.

In the first phase, the model performs open information extraction, transforming text into semantic triplets where any noun phrase can become a subject or object—the model is not constrained to a pre-extracted entity list. Unlike closed information extraction, which requires output triplets to follow a pre-defined schema, open extraction enables capture of domain-specific vocabulary (Zhang and Soh, 2024).

In the second phase, the model generates natural language definitions for each predicate, forcing articulation of what each relation means. This definition phase is crucial: LLMs can justify their extractions via interpretable explanations, and the definitions guide canonicalization by identifying the closest entity and relation type candidates (Zhang and Soh, 2024). For example, the predicate *escribió* receives a definition specifying that the subject produced or composed the object (text, work, etc.). In the third phase, embedding-based retrieval matches open predicates to a target schema, with LLM verification for ambiguous cases.

This open extraction approach aims to preserve the semantic texture of the source text, including discipline-specific vocabulary and Spanish-language predicates. The design produces a higher proportion of abstract concepts (42.8% of top entities), Spanish predicates preserved in their original form, and 21,025 semantic definitions. From 4,728 chunks, EDC extracted 106,883 raw triplets.

3.4. Architectural Comparison

The two approaches represent fundamentally different extraction philosophies that reflect broader tensions in knowledge graph construction research. KGGen discovers entities first and constrains relations to those entities, normalizes predicates to English, generates no semantic definitions, and employs a multi-stage pipeline with entity resolution as a post-processing step (Gemini 2.0 Flash; Gemini Team, 2024). Its design prioritizes graph connectivity and relation reusability, addressing the sparsity problem described above. EDC, by contrast, allows entities to emerge from triplet extraction without prior constraints, preserves source-language predicates (Spanish), generates explicit definitions for each predicate, and employs a three-phase pipeline with explicit schema canonicalization (implemented here with Mistral; Jiang et al., 2023). Its design prioritizes semantic transparency: by generating definitions, the system makes predicate semantics explicit and enables disambiguation of polysemous terms (Zhang and Soh, 2024).

These architectural differences have consequences for what each system captures. The entity-first approach produces graphs biased toward proper names—authors, places, institutions—that are unambiguously “entities” in the NER sense. The open approach captures more conceptual vocabulary—theoretical terms, aesthetic categories—at the cost of noisier entity boundaries. The choice between implicit schema (embedded in LLM weights and post-hoc entity resolution) and explicit schema (generated definitions plus canonicalization) also affects interpretability: EDC’s definitions make predicate semantics transparent, while KGGen’s normalized predicates facilitate cross-corpus comparison but obscure source-language nuances. Both approaches represent advances over earlier extraction methods that, as reviewed in the literature (Kamp et al., 2023; Ye et al., 2023), either required pre-defined schemas in the prompt—limiting scalability to complex domains—or produced uncanonicalized open knowledge graphs with substantial redundancy. Systematic comparisons of extraction pipelines for knowledge graphs remain rare (Jaradeh et al., 2023), and no prior work has compared these specific architectures on humanities text.

Our comparison should be read with the awareness that KGGen is implemented here with Gemini 2.0 Flash and EDC with Mistral, so observed differences may partly reflect the underlying LLMs rather than extraction architecture alone. This confound, however, reinforces rather than weakens convergent findings, since they hold despite differences in both architecture and LLM.

3.5. Entity Canonicalization for Comparative Analysis

Raw extraction outputs differ substantially in scale: KGGen produced 104,500 triplets with 83,344 unique entities and 19,508 unique predicates, while EDC produced 106,883 triplets with 88,803 unique entities and 22,328 unique predicates. Comparing raw graphs directly would confound architectural differences with scale differences and unresolved duplicates.

To construct a shared entity set for comparison, we applied a curation strategy anchored on KGGen's output. Manual entity curation was necessary because neither method fully resolves entity duplicates: KGGen's resolution failed on Spanish name variants (as discussed above), and EDC by design addresses only predicate canonicalization, not entity de-duplication (Zhang and Soh, 2024). For KGGen, we performed manual review of entities with degree $\geq$ 100, yielding 106 canonical entities with type classifications. For EDC, we used fuzzy string matching to map raw entities to KGGen's 106 canonical entities, using minimum match length of 4 characters and similarity threshold of 0.85. The degree threshold was chosen to focus on major disciplinary patterns—prominent authors, core concepts, key institutions—that can be verified against domain expertise in Argentine literary studies. All comparative analyses use these curated graphs with matched entity sets.

The 106 curated entities were classified into a domain-specific taxonomy reflecting the structure of literary scholarship: persons (44), including literary authors, critics, and theorists; concepts (25), divided into cultural terms (*cultura*, *historia*) and literary terms (*escritura*, *poética*); places (15), from countries to cities; organizations (9), including universities, journals, and publishers; and textual roles (13) such as *autor* 'author', *lector* 'reader', and *narrador* 'narrator'. This taxonomy captures both the object of study (authors, works, places) and the meta-discourse of criticism (concepts, roles).

Both extraction pipelines embed provenance metadata at extraction time: article identifier (which of 472 articles the triplet comes from), publication year (1996–2024), topic cluster (BERTopic assignment), and text chunk. This enables filtering and analysis by source, time period, and topic—essential for scholarly applications where citation and traceability matter.

4. Results

We present comparative results focusing on extraction statistics, entity composition, convergent findings that hold across methods, and divergent findings that reveal architectural differences.

4.1. Extraction Statistics

Both methods processed identical input (4,728 text chunks from 472 articles) and produced comparable raw output volumes, as shown in Table 3. EDC produces slightly more triplets (+2.3%) and substantially more unique predicates (+14.5%), consistent with its open extraction architecture. The 21,025 semantic definitions generated by EDC's definition phase have no equivalent in KGGen.

After curation to the same 106 canonical entities (Table 4), KGGen produced 2,093 edges while EDC produced 2,412 edges (+15%). Two entities had no EDC matches due to naming variations.

Metric	KGGen	EDC	Diff.
Raw triplets	104,500	106,883	+2.3%
Unique entities	83,344	88,803	+6.5%
Unique predicates	19,508	22,328	+14.5%
Semantic definitions	0	21,025	—

Table 3: Raw extraction output comparison.

Metric	KGGen	EDC
Curated entities	106	104*
Curated edges	2,093	2,412
Edge difference	—	+15%

*Two entities had no EDC matches.

Table 4: Curated graph comparison with matched entities.

4.2. Entity Composition

The two methods capture different entity distributions, as shown in Table 5. Analyzing the top 100 entities by frequency reveals substantial differences in what each method prioritizes.

Entity Type	KGGen	EDC
Proper names (authors, places)	40.7%	18.7%
Generic concepts	31.2%	42.8%
Years/identifiers	5.1%	8.2%
Abstract nouns	23.0%	30.3%

Table 5: Entity type distribution in top 100 entities.

KGGen's entity-first constraint biases extraction toward proper names—entities that are unambiguously recognizable. EDC's open architecture captures more abstract concepts and generic noun phrases. This reflects the architectural difference: constraining relations to pre-extracted entities favors named entity recognition patterns over conceptual vocabulary extraction.

4.3. Convergent Finding: Cultural vs. Textual Framing

We tested a hypothesis grounded in critical historiography of Argentine literary studies. Since its emergence in the 1950s, literary criticism in Argentina has been defined by an orientation toward "literature's outside"—following Blanchot's concept of literature's tension with culture (Blanchot, 1971), which Argentine criticism has historically operationalized as the practice of addressing literature through its social, cultural, and political dimensions (Cortés, 2021; Dalmaroni, 2004). The influential journal *Contorno* (1953–1959) inaugurated this orientation by confronting the aestheticism of *Sur* magazine and linking literature to historical events, influenced by Jean-Paul Sartre's theory of committed literature (Cella, 1999). This orientation persisted through subsequent decades: Dalmaroni (2004) traces a shift from "literature and society" in the 1950s to "culture and politics" in the 1980s and 1990s, shaped by the reception of British cultural studies in Argentina through journals like *Punto de vista* (1978–2008) (Sarlo, 2000; Richard, 2001).

Our hypothesis is that this orientation toward literature's outside—now articulated through cultural and political vocabulary—should be detectable in the knowledge graphs extracted from Orbis Tertius, from a time span of almost 30 years. We operationalized this by measuring edge weights between *literatura* 'literature' and two concept clusters:

Cultural concepts: *cultura* 'culture', *historia* 'history', *sociedad* 'society', *política* 'politics', *arte* 'art', *memoria* 'memory', *nación* 'nation', *mundo* 'world'

Textual concepts: *texto* 'text', *escritura* 'writing', *lectura* 'reading', *lenguaje* 'language', *palabra* 'word', *forma* 'form', *narración* 'narration'

Method	Cultural	Textual	Ratio
KGGen	101	40	2.52×
EDC	126	58	2.17×

Table 6: Cultural vs. textual framing weights.

Both methods show cultural framing dominates by approximately 2.2–2.5×. Chi-squared goodness-of-fit tests against the null hypothesis of equal weights confirm the imbalance is statistically significant for both KGGen ($\chi^2 = 26.4$, $p < .001$) and EDC ($\chi^2 = 25.1$, $p < .001$, df $= 1$). The totals are driven by a subset of each category: 4 of 8 cultural concepts (*cultura*, *historia*, *arte*, *sociedad*) and 4 of 7 textual concepts (*texto*, *escritura*, *lectura*, *lenguaje*) showed non-zero connections to *literatura*; the remaining concepts (*política*, *memoria*, *nación*, *mundo*, *palabra*, *forma*, *narración*) contributed zero edges in the KGGen curated graph. This convergence across fundamentally different extraction architectures strengthens confidence that the finding reflects textual and semantic patterns in the corpus. Argentine literary studies in Orbis Tertius frames literature primarily as a cultural phenomenon—embedded in social, historical, and political contexts—rather than as a purely textual object for formalist analysis.

To summarize, the quantitative pattern detected by both extraction methods aligns with the critical historiography outlined above: the orientation toward literature's outside that Dalmaroni (2004) traces from *Contorno* through *Punto de vista* persists in three decades of Orbis Tertius scholarship. The convergence across architecturally different methods—despite different entity constraints, predicate languages, and processing pipelines—reinforces our hypothesis that this cultural framing reflects a genuine discursive orientation in the corpus.

4.4. Convergent Finding: Author Networks

Both methods identify consistent author-to-author connections in the curated graphs. The scholarly consensus on key relationships emerges regardless of extraction architecture. Several connections show high edge weights in both methods: Borges–Groussac (reflecting historical succession in Argentine letters), Borges–Cortázar (the canonical pairing of mid-century fiction), Piglia–Arlt (Ricardo Piglia's critical recovery of Roberto Arlt as a major figure), and Darío–Groussac (the modernismo network connecting Rubén Darío to Argentine intellectuals). Contemporary connections like Aira–Piglia show medium weights in both methods, reflecting ongoing critical dialogue.

However, while the same author pairs emerge from both methods, their relative weights differ substantially. A Spearman rank correlation on the 12 pairs shared between both methods' top connections yields $\rho = 0.12$ ($p > .05$, $n = 12$), indicating weak rank concordance. For instance, Borges–Cortázar ranks among KGGen's strongest connections (weight 19) but appears with low weight in EDC (weight 3), while the Ocampo–Ocampo connection shows the reverse pattern (KGGen rank 9, EDC rank 2). This suggests that while both methods identify the same scholarly community structure, edge weights reflect extraction architecture as much as corpus signal.

The Borges-centrality finding is particularly robust: Borges appears as the highest-degree author node in both curated graphs, connected to authors across periods (Lugones, Groussac, Cortázar, Piglia) and to both cultural and textual concept clusters. This convergence across architecturally different extraction methods strengthens confidence that the centrality of Borges in Argentine literary studies—a well-established claim in critical historiography—is indeed reflected in the discursive patterns of Orbis Tertius scholarship.

4.5. Divergent Finding: Predicate Characteristics

The methods diverge significantly in predicate extraction, as summarized in Table 7.

Aspect	KGGen	EDC
Predicate language	English	Spanish
Unique predicates (curated)	2,499	4,891
Analysis predicates	3.7%	5.9%
Semantic definitions	0	21,025

Table 7: Predicate characteristics comparison.

KGGen normalizes predicates to English regardless of source language, while EDC preserves Spanish predicates. EDC captures nearly twice as many unique predicates after curation, and shows higher proportion of scholarly analysis predicates—*analiza* 'analyzes', *estudia* 'studies', *critica* 'critiques', *interpreta* 'interprets'—suggesting it better captures the meta-discourse of literary studies rather than just the object discourse.

EDC's definition phase produces explicit meanings that enable semantic disambiguation, as illustrated in Table 8.

Predicate	EDC Definition
escribió	The subject produced or composed the object (text, work, etc.)
influyó	The subject exercised intellectual, aesthetic, or literary influence over the object
dialoga con	The subject establishes an intertextual or conceptual relationship with the object
reescribe	The subject transforms or reworks elements of the object in their own work

Table 8: Example semantic definitions generated by EDC (translated from Spanish).

Such definitions enable distinguishing authorship from analysis that KGGen's normalized predicates cannot capture. However, KGGen's normalized English predicates facilitate cross-corpus and cross-lingual comparison, while EDC's Spanish predicates preserve linguistic specificity important for humanities scholarship where terminology carries disciplinary weight.

4.6. Summary

Findings that converge across methods—the cultural hypothesis ratio and core author networks—can be reported with confidence as reflecting corpus patterns. Method-dependent findings—absolute counts, predicate language, entity composition—require acknowledging the extraction method when interpreting results.

5. Discussion

The comparative analysis reveals that extraction architecture shapes what gets captured in knowledge graphs from scholarly text. KGGen's entity-first constraint ensures relations are anchored on recognizable entities, producing cleaner graphs with higher proportion of proper names that map well to traditional NER categories. Its predicate normalization facilitates cross-corpus comparison, and its two-phase architecture is simpler and more predictable. EDC preserves semantic texture through Spanish predicates and explicit definitions—important for humanities scholarship where terminology carries disciplinary weight. Its open extraction architecture captures more abstract concepts and theoretical terms that NER-based approaches miss, and its higher proportion of analysis predicates indicates it captures not just what critics discuss but how they discuss it.

Our entity curation was anchored on KGGen's output: entities with degree $\geq$ 100 were selected from KGGen, and EDC's output was then fuzzy-matched to this set. This design prioritizes entities prominent enough to represent major disciplinary patterns verifiable by domain experts. A consequence is that EDC's unique entities—those that KGGen did not extract with sufficient frequency—are not represented in the comparison. However, all 20 of EDC's highest-degree entities appear among the 106 KGGen-curated entities, suggesting that both methods converge on which entities are most prominent. Of KGGen's 104,500 raw triplets, only 7,025 (6.7%) involved curated entities. Our comparative analysis therefore characterizes the densest core of the knowledge graph—the most prominent entities and connections—rather than the full extracted graph.

To test whether this curation choice drives the convergent findings, we performed a reverse analysis anchored on EDC's output, selecting EDC's 115 highest-degree entities (degree $\geq$ 100) independently of KGGen; full results are available in the accompanying repository. Of these, 82 have counterparts in KGGen's curated set via fuzzy string matching, while 33 are unique to EDC—predominantly generic literary terms (*artículo*, *protagonista*, *relato*, *revista*) and domain-specific concepts absent from the curated set (*parodia*, *surrealismo*, *exilio*). In the EDC-anchored curated graph, author networks and Borges centrality are confirmed: the same author pairs emerge (Darío–Groussac, Ocampo–Borges, Aira–Piglia, Borges–Cortázar), and Borges remains the highest-degree author. The cultural framing analysis requires attention to the degree threshold. The threshold of 100, chosen for KGGen's degree distribution, is not equally calibrated to EDC's: cultural concepts like *cultura* (degree 98), *sociedad* (52), and *política* (60) fall just below it. At degree ≥ 50, which includes the full cultural vocabulary, the ratio is $2.71\times$, consistent with the KGGen-anchored results. The core convergent findings thus hold regardless of which method anchors the curation, though the appropriate degree threshold varies with extraction architecture.

The comparison suggests a methodological principle grounded in methodological triangulation (Heesen et al., 2019): findings that converge across architecturally different extraction methods merit higher confidence than findings from a single method, even when neither method can be independently validated. Our convergent findings held despite different entity constraints, predicate languages, and processing pipelines. This convergence does not validate individual triplets, but it provides stronger evidence that aggregate patterns exist in the corpus rather than being artifacts of a particular extraction architecture. Conversely, method-dependent findings should be reported with appropriate caveats. Researchers using knowledge graphs for literary analysis should avoid overinterpreting patterns that might reflect extraction architecture rather than corpus characteristics.

For hypothesis testing, we recommend using multiple extraction methods when possible; if a finding holds across methods, confidence increases. For author and entity networks, either method produces consistent core networks—choose based on downstream needs. For conceptual vocabulary, EDC's open architecture is more suitable when research questions involve how scholars use concepts rather than just which authors they discuss. For cross-lingual work, KGGen's English normalization facilitates comparison across languages. For Spanish-language scholarship, EDC's preservation of source-language predicates respects linguistic specificity.

Knowledge graph construction for humanities scholarship requires attention to what gets lost in extraction. Both methods produce useful graphs, but neither fully captures the interpretive texture of literary studies. The cultural framing finding illustrates what knowledge graphs can capture: aggregate patterns in how a scholarly community discusses literature over time. The predicate limitations illustrate what they cannot capture: the force of individual interpretations, the rhetorical moves of close reading. Knowledge graphs from humanities scholarship are best understood as maps of discourse topology—who discusses whom, what concepts co-occur—which constitutes one dimension of literary knowledge that combines with interpretive argumentation, rhetorical strategy, and historical situatedness to produce scholarly meaning.

6. Limitations

Knowledge graph extraction from humanities scholarship lacks established ground truth. Even for general open information extraction, "a clear and formal specification of what a valid relational tuple consists of is still missing" (Kamp et al., 2023); this challenge is amplified in interpretive domains like literary studies, where "correct" extraction is underdefined. We rely on convergence across methods as a form of methodological triangulation to identify robust aggregate patterns, though this does not constitute gold-standard evaluation and cannot validate individual extracted triplets. The field needs annotated corpora for humanities KG extraction—a resource that does not currently exist for literary studies in Spanish.

Knowledge graphs capture the topology of scholarly discourse but not the force of interpretation. Consider two articles that both link "Borges" to "gauchesca": one might argue Borges parodies the genre, another might argue he continues it. Both produce the same triplet, but the interpretive content differs fundamentally. This limitation is structural: knowledge graphs represent what is said but not the hermeneutic force of the saying.

Our specific comparison introduces a further limitation. KGGen (Gemini 2.0 Flash) and EDC (Mistral) use different underlying LLMs, so divergent findings—entity composition, predicate vocabulary—may reflect the underlying models as much as extraction architecture. Convergent findings are affected in the opposite direction: agreement despite different LLMs strengthens the evidence that a pattern originates in the corpus rather than in either model.

Both extraction methods rely on LLMs trained predominantly on English-language data. This may bias entity recognition toward figures and concepts prominent in anglophone scholarship, and could disadvantage domain-specific Spanish vocabulary that appears infrequently in training corpora. KGGen's normalization of predicates to English makes this bias explicit; EDC's preservation of Spanish predicates mitigates it at the predicate level but not at the entity level.

Our findings derive from a single journal from a specific institution. However, with more than 20 active peer-reviewed literary studies journals in Argentina, the methodology presented here is readily scalable to broader comparative analyses across the field. We compare methods on Spanish text only; the comparison between KGGen's English normalization and EDC's Spanish preservation would differ for English-language corpora. Results depend on our degree threshold for entity curation, and both methods depend on specific LLM versions that may behave differently with updates.

7. Conclusion

We compared two LLM-based knowledge graph extraction approaches—entity-anchored (KGGen) and open extraction (EDC)—on 472 Spanish-language literary studies articles from Orbis Tertius (1996–2024). Despite fundamental architectural differences, both methods converge on key findings: cultural framing dominates literary discourse by 2.2–2.5× over textual framing, and core author networks with Borges as central figure are consistent across extraction approaches. This convergence strengthens confidence that these findings reflect the corpus rather than method artifacts, and demonstrates the utility of knowledge graphs for analyzing academic texts from well-defined disciplines such as literary studies, where language and meaning intertwine in complex ways.

The methods diverge in what they capture. KGGen's entity-first constraint produces graphs biased toward proper names (40.7% vs. 18.7%), with normalized English predicates suitable for cross-corpus comparison. EDC's open architecture captures more conceptual vocabulary (42.8% abstract concepts), preserves Spanish predicates, and generates semantic definitions that enable predicate disambiguation.

Based on our comparison, we suggest four practices for knowledge graph construction from scholarly text: test robustness by comparing extraction methods when feasible; select entity-anchored approaches for cross-corpus comparison and open extraction for capturing conceptual vocabulary; distinguish method-dependent from method-independent findings; and treat discourse topology as one dimension of scholarly knowledge rather than a complete representation of it. The cultural framing finding—that Argentine literary studies frames literature primarily through social and historical contexts—aligns with documented influence of cultural studies on Latin American scholarship and holds across methods, suggesting it reflects genuine discursive patterns in three decades of published literary criticism.

This study demonstrates the potential of natural language processing methods for literary studies, offering knowledge graphs as tools for exploration, hypothesis testing, and large-scale pattern discovery that complement traditional close reading approaches. An interactive visualization of the curated graph and the complete entity taxonomy are available in the accompanying repository.[1]

Acknowledgments

This work was supported by a Georg Forster Research Fellowship from the Alexander von Humboldt Foundation.

Ethical Considerations

This study analyzes publicly available open-access academic articles. No human subjects were involved. The corpus is freely accessible through the journal's OJS platform.

8. Bibliographical References

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 344–354, Beijing, China. Association for Computational Linguistics.

Fernanda Beigel and Maximiliano Salatino. 2015. Circuitos segmentados de consagración académica: las revistas de ciencias sociales y humanas en la argentina. *Información, cultura y sociedad*, (32).

Haonan Bian. 2025. LLM-empowered knowledge graph construction: A survey.

Maurice Blanchot. 1971. *L'amitié*. Gallimard.

Susana Cella. 1999. *Historia crítica de la literatura argentina vol. 10*. Emecé Editores.

Federico Gabriel Cortés. 2021. La palabra errante en maurice blanchot: Literatura como impugnación. *Landa*, (10(1)).

Miguel Dalmaroni. 2004. *La palabra justa: Literatura, crítica y memoria en la Argentina, 1960-2002*. Melusina.

Gemini Team. 2024. Gemini: A family of highly capable multimodal models.

Analía Gerbaudo. 2024. *Tanto con tan poco: los estudios literarios en Argentina 1958-2015*. Ediciones UNL.

[1] Data and code: https://github.com/fedexx1/orbis-kg

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure.

Remco Heesen, Liam Kofi Bright, and Andrew Zucker. 2019. Vindicating methodological triangulation. *Synthese*, 196(8):3067–3081.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Mohamad Yaser Jaradeh, Kuldeep Singh, Markus Stocker, Andreas Both, and Sören Auer. 2023. Information extraction pipelines for knowledge graphs. *Knowledge and Information Systems*, 65(5):1989–2016.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Sreejith Sudhir Kalathil, Tian Li, Hang Dong, and Huizhi Liang. 2024. HTEKG: A human-trait-enhanced literary knowledge graph with language model evaluation. In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024)*, pages 207–215. SCITEPRESS.

Serafina Kamp, Morteza Fayazi, Zineb Benameur-El, Shuyan Yu, and Ronald Dreslinski. 2023. Open information extraction: A review of baseline techniques, approaches, and applications.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2023. Dspy: Compiling declarative language model calls into self-improving pipelines.

Jiayin Lan, Jiaqi Li, Baoxin Wang, Ming Liu, Dayong Wu, Shijin Wang, and Bing Qin. 2025. NLP-AKG: Few-shot construction of NLP academic knowledge graph based on LLM.

Bo Li, Gexiang Fang, Yang Yang, Quansen Wang, Wei Ye, Wen Zhao, and Shikun Zhang. 2023. Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness.

Belinda Mo, Kyssen Yu, Joshua Kazdan, Joan Cabezas, Proud Mpala, Lisa Yu, Chris Cundy, Charilaos Kanatsoulis, and Sanmi Koyejo. 2025. Kggen: Extracting knowledge graphs from plain text with language models.

Roxana Patiño. 1997. Intelectuales en transición: las revistas culturales argentinas (1981- 1987). *Cuadernos de Recienvenido*, (4).

Nelly Richard. 2001. Globalización académica, estudios culturales y crítica latinoamericana.

Beatriz Sarlo. 2000. Los estudios culturales y la crítica literaria en la encrucijada valorativa. *Revista de Crítica Cultural*, 21:32–38.

Marco Antonio Stranisci, Eleonora Bernasconi, Viviana Patti, Stefano Ferilli, Miguel Ceriani, and Rossana Damiano. 2023. The world literature knowledge graph. In *The Semantic Web – ISWC 2023*, volume 14266 of *Lecture Notes in Computer Science*, pages 435–452. Springer.

Carolina Unzurrunzaga, Cecilia Rozemblum, Cristian Pucacco, Gustavo Parente, and Mauricio Esterellas. 2015. Ojs implementation and development of the scientific journals site of the school of humanities and education sciences at the universidad nacional de la plata. *Scholarly and Research Communication*, (6(1)).

Shikhar Vashishth, Prince Jain, and Partha Talukdar. 2018. Cesi: Canonicalizing open knowledge bases using embeddings and side information. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, page 1317–1327. ACM Press.

Jiannan Wang, Tim Kraska, Michael J. Franklin, and Jianhua Feng. 2012. Crowder: Crowdsourcing entity resolution.

Hongbin Ye, Ningyu Zhang, Hui Chen, and Huajun Chen. 2023. Generative knowledge graph construction: A review.

Bowen Zhang and Harold Soh. 2024. Extract, define, canonicalize: An llm-based framework for knowledge graph construction.