



MEANING GRAPHS FOR THE HUMANITIES IN THE AGE OF LLMs: AMR AND UMR AS INTERPRETIVE INFRASTRUCTURE

CHONGWON PARK *

Abstract Large language models (LLMs) have made language technologies widely accessible to humanities scholars, but they have also intensified concerns about transparency, reproducibility, and interpretive responsibility. This article argues that graph-based meaning representations, especially Abstract Meaning Representation (AMR) and Uniform Meaning Representation (UMR), can function as interpretive infrastructure for humanities research in the age of LLMs. Meaning graphs are “thin” by design in that they deliberately encode a constrained set of semantic distinctions. That selectivity is not a weakness for humanistic inquiry; rather, it enables a disciplined workflow in which researchers can separate (i) the semantic commitments that a text licenses (events, participants, temporal and modal dependencies) from (ii) richer interpretive claims (stance, ideology, affect, narrative framing) that can be layered on top. I review AMR and UMR at a level accessible to humanities audiences, discuss what changes in the LLM era (including both opportunities and limits of using LLMs for semantic parsing), and propose humanities-centered workflows and research questions. Several compact sample analyses illustrate how meaning graphs can support interpretive tasks in historiography, narrative analysis, and translation studies. A final section explicitly lists resources (datasets, tools, and guidelines) to support reproducible experimentation.

Keywords: Abstract Meaning Representation (AMR), Uniform Meaning Representation (UMR), meaning graphs, digital humanities, large language models

1. INTRODUCTION: HUMANITIES RESEARCH UNDER GENERATIVE AI

Generative LLMs have changed the interface between scholars and computation. Tasks that once required specialized NLP pipelines, such as summarizing corpora, extracting motifs, clustering themes, or classifying passages by topic, can now be attempted with prompts. This shift has two consequences that matter for the humanities.

The first consequence is practical. LLMs can reduce the cost of exploratory analysis and lower

* University of Minnesota Duluth

the barriers to working with large archives. Recent work in digital humanities documents both enthusiasm and caution. Researchers describe LLMs as a “colleague” or a “toolbox,” but they also express anxiety about accuracy, authorship, labor, and disciplinary identity (Ma et al. 2024). Workflow-oriented discussions emphasize the importance of documenting how generative AI is used in research practice, rather than treating model outputs as self-evident results (Armaselu 2024). Across these discussions, the recurrent theme is not only whether the technology is useful, but whether its use can be made accountable. The second consequence is epistemic. Prompt-based interaction can obscure evidential chains. A model may produce a fluent summary or interpretation without exposing what in the source text supports each claim. That opacity collides with core humanistic norms: argumentation, citation, and responsible interpretation. Historians have therefore argued that LLMs should be treated as sources that require new forms of source criticism, because their capabilities and blind spots are shaped by the uneven distribution of digitized cultural memory (Hutchinson 2024). More broadly, calls for reproducibility and explainability in digital humanities and adjacent fields have intensified (Ries et al. 2024).

This paper offers a proposal grounded in meaning representation. I argue that explicit meaning graphs, particularly AMR and UMR, offer a practical way to make semantic commitments visible, comparable, and discussable. They do not replace close reading.¹ Instead, they provide a structured “mid-level” artifact between raw text and higher-level interpretation. Such artifacts can help scholars (i) document what kinds of semantic claims an analysis depends on, (ii) compare those claims across texts, translations, or genres, and (iii) use LLMs more responsibly by separating model-generated suggestions from audited semantic structure. Two design features make these representations especially relevant. First, both represent meaning as graphs, which supports qualitative inspection as well as quantitative querying. Second, UMR is explicitly designed for cross-linguistic and document-level semantics, two concerns central to humanities fields such as comparative literature, translation studies, and historiography (van Gysel et al. 2021; Bonn et al. 2024).

This is a humanities paper. I assume no prior familiarity with semantic parsing, but I avoid treating AMR/UMR as merely technical tools. My focus is methodological. I connect meaning graphs to interpretive practices, propose workflows and research questions, and illustrate what “graph-grounded” analysis can look like with compact examples. The examples are intentionally schematic. They show the kind of reasoning the representations enable, not a complete annotation specification for the humanities.

I will provide a roadmap for the argument by first motivating the need for auditable intermediate representations in humanities work with LLMs (Section 3). I then introduce AMR and UMR with concrete examples (Section 4), discuss what changes in the LLM era (Section 5), present sample analyses that illustrate graph-grounded reading (Section 6), and outline a humanities-centered

¹ AMR and UMR are graph-based meaning representations that encode who did what to whom, when, and under what modal or discourse conditions, abstracting away from surface syntax. A meaning representation makes semantic commitments explicit by forcing choices about predicate identity, argument structure, scope, and discourse relations, rather than leaving them implicit in wording alone.

research agenda (Section 7). I also include an explicit discussion of datasets, tools, and other resources that make this approach feasible in practice (Section 9).

2. LLMS AS HUMANITIES TOOLS AND OBJECTS

The theme of “humanities in the age of AI” is often framed as a question of adoption: how should scholars use new tools? A complementary framing treats AI as an object of humanities inquiry. Both framings matter for methodological design.

2.1 LLMs as Tools in Digital Humanities Practice

Recent studies of digital humanities researchers describe LLMs as simultaneously attractive and unsettling. They promise speed, breadth, and an accessible interface, yet they raise questions about methodological legitimacy and scholarly voice (Ma et al. 2024). Workflow-oriented discussions therefore emphasize documentation, such as what model was used, with what prompts, on what data, and with what validation (Armaselu 2024). Such documentation is not bureaucratic overhead. It is the computational analogue of citation practices and archival description.

The range of uses is broad. Humanities researchers increasingly explore LLMs for tasks such as exploratory summarization, candidate topic generation, translation assistance, extraction of named entities and events, and support for qualitative coding. Accounts of how generative AI is changing digital humanities research and services highlight both opportunities (new modes of interaction with collections) and risks (hallucination, homogenization of style, and hidden bias) (Liu et al. 2024). In addition, multimodal LLMs are being tested as annotators of cultural artifacts that are not purely textual, suggesting that the boundary between “text analysis” and “artifact analysis” may shift as models become more general-purpose (Suviranta and Hiippala 2025).

2.2 LLMs as Cultural Technologies and Sources

At the same time, LLMs are not neutral tools. They embody particular histories of digitization, data curation, and institutional power. Humanities scholars therefore increasingly treat LLM behavior as a cultural phenomenon to be interpreted and criticized, not merely optimized. Work on computational hermeneutics explicitly frames generative AI as a cultural technology whose evaluation must include questions of mediation, labor, and evidential standards (Kommers et al. 2026). In a historiographic register, LLMs have been analyzed as objects requiring source criticism, because what they “know” is shaped by what has been digitized, translated, and made machine-readable (Hutchinson 2024). This aligns with a broader humanities concern. Modern knowledge infrastructures are never transparent, and scholarly responsibility includes analyzing their conditions of possibility.

The humanities also contribute by asking distinctive questions *about* LLMs. For example, recent

work in *Digital Scholarship in the Humanities* investigates semiotic and perceptual phenomena such as iconicity in LLM outputs, showing how humanities-oriented concepts can be operationalized as empirical questions about generative systems (Marklová et al. 2025). These lines of work emphasize that the humanities are not only end-users of AI. Rather, they are critical analysts of how AI reorganizes meaning-making.

2.3 Why Meaning Representations Help in Both Roles

Treating LLMs as both tools and objects highlights a shared methodological need: representations that make the mediation visible. If LLMs are used to produce scholarly claims, we need ways to trace those claims to evidence. If LLMs are studied as cultural objects, we need ways to characterize what kinds of meanings they preserve, distort, or invent.

Meaning graphs offer a practical bridge. They force explicit decisions about predicate–argument structure, reference, temporality, and attribution.² Those decisions can be debated in familiar humanities terms (voice, viewpoint, narrative sequence, evidential stance), while also being operationalized for corpus-scale comparison. In other words, meaning graphs provide an interface where humanities critique (about mediation and interpretation) can meet computational practice (about representation and analysis).

3. WHY EXPLICIT MEANING IS A HUMANITIES PROBLEM

Concerns about hallucination are real, but they are not the whole story. The deeper humanities concern is *auditability*. When a model produces an answer, what does the answer actually commit to, and how can a scholar trace those commitments back to evidence?

3.1 From Hallucination to Epistemic Opacity

Hallucinations illustrate a simple failure mode. Models sometimes produce false statements. Yet even when a model output is broadly “correct,” it may still be epistemically opaque. A summary may subtly shift attribution (turning a reported claim into an asserted fact), a paraphrase may alter agency (adding or erasing actors), and a thematic analysis may conflate textual evidence with background generalizations. These are familiar worries in historiography and literary studies. They resemble problems of paraphrase, translation, and interpretation. What is new is that LLMs can perform such transformations at scale and with persuasive fluency.

² Meaning representations such as AMR and UMR are not semantic theories in the formal sense, unlike frameworks such as Neo-Davidsonian event semantics. They do not aim to explain the nature of meaning or its compositional derivation. Instead, they function as representational conventions that make semantic decisions explicit and inspectable. In this sense, they operationalize semantics by requiring concrete commitments about events, participants, reference, temporal structure, and attribution, even though they do not provide a full explanatory theory of meaning itself.

This is why many humanities-facing discussions focus on method rather than mere error rates. Computational hermeneutics frames generative AI as a cultural technology whose use must be evaluated by how it reorganizes interpretive labor and evidential norms (Kommers et al. 2026). Similarly, work on qualitative coding with LLMs emphasizes that scaling interpretation requires documentation of categories, prompts, and validation procedures, not simply producing labels (Dunivin 2025). Digital humanities discussions of reproducibility and explainability make the same point in a different register. Results are only scholarly if others can inspect, reproduce, and critique the chain from sources to claims (Ries et al. 2024).

3.2 Meaning Graphs as Auditable Mediation

Meaning representations provide one concrete response. They create explicit objects that sit between text and interpretation. A meaning graph is a map, not the territory. It is a selective encoding of semantic commitments that annotators can agree on with reasonable reliability. The idea is not to replace interpretation, but to stabilize some of the inferential moves that interpretation relies on. I propose the term *graph-grounded reading* for workflows in which meaning graphs serve as scaffolds for interpretive work. The approach has three commitments. It treats the graph as a thin semantic core that encodes a constrained set of meaning commitments, including events, participants, core semantic relations, and, in UMR, document-level links. It represents humanities-relevant categories such as stance, affect, rhetorical framing, and narrative roles as additional layers linked to graph nodes or subgraphs rather than forcing them into the core. It also insists on an auditable role for LLMs. Models may propose candidate graphs or candidate interpretive labels, but the resulting structures are treated as scholarly artifacts that can be cited, revised, and critiqued.

The word “thin” may sound like a concession, but it is strategic. A meaning graph is designed to be readable and annotatable. Its methodological value lies in enabling scholars to say, “This interpretive claim presupposes that the text asserts event e with participant x as agent, and that e temporally follows event e' .” Once such commitments are explicit, disagreement can be localized. The dispute may be about a graph decision, about an interpretive layer, or about the mapping between them.

4. AMR AND UMR: WHAT THEY ENCODE

This section introduces AMR and UMR for a humanities audience. I emphasize not only what the frameworks represent, but how their design choices align with humanities concerns.

4.1 AMR in a Nutshell

AMR was introduced as a whole-sentence meaning representation with the explicit goal of building a large, consistent “meaning bank” (Banarescu et al. 2013). An AMR is typically a rooted, directed graph. Nodes correspond to concepts (events, entities, properties, quantities), and edges correspond to labeled relations between those concepts. Many predicates use PROPBANK framesets (Palmer et al. 2005), which support stability across syntactic alternations.³

The most common textual encoding is PENMAN notation (Goodman 2020). PENMAN provides a compact parenthesized syntax with variables for nodes.⁴ A variable can be reused to express reentrancy, the mechanism for within-sentence coreference. Example (1) and Figure 1 provide a first illustration.

- (1) *The student protested the policy.*

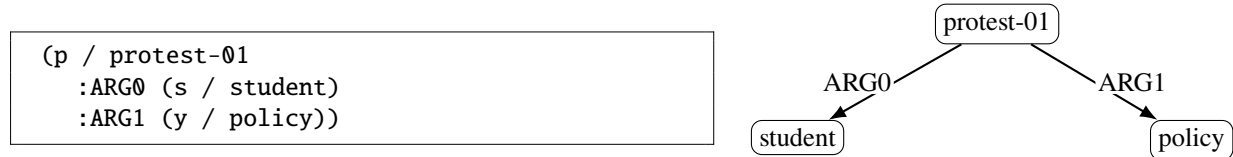


Figure 1 A basic AMR in PENMAN notation (left) and a graph view (right) for (1).

Two properties of AMR are especially relevant for humanities use. First, AMR is designed to be relatively invariant to changes in surface syntax. Second, AMR includes conventions for frequent phenomena (named entities, negation, quantities, within-sentence coreference) that make the graphs systematically searchable.

AMR is designed to be relatively stable across syntactic alternations. Examples (2) and (3) differ in voice, but they typically map to the same core AMR structure, as shown in Figure 2.

- (2) *The committee approved the proposal.*

- (3) *The proposal was approved by the committee.*

³ PROPBANK framesets define predicate senses together with their core argument roles, abstracting away from surface syntactic realizations such as active and passive voice. By assigning consistent roles (e.g., ARG0, ARG1) across alternations, framesets support stable predicate meaning and comparability across different syntactic forms.

⁴ PENMAN notation is a linear, parenthesized way of writing a graph as text. Each pair of parentheses introduces a node, variables name nodes so they can be referred to more than once, and labeled edges indicate semantic relations between nodes. This notation makes it possible to represent graph structure in plain text while preserving node identity and relational structure.

```
(a / approve-01
  :ARG0 (c / committee)
  :ARG1 (p / proposal))
```

Figure 2 Shared AMR core (PENMAN notation) for (2) and (3).

For many NLP tasks, collapsing active and passive is desirable. For humanities analysis, the collapse is not always desirable, because voice can be rhetorically meaningful. The graph nonetheless offers a controlled separation. The semantic core, i.e., who did what to whom, can be compared across texts, while voice and other stylistic choices can be studied as separate layers. AMR also uses a set of high-frequency conventions that make certain recurring distinctions explicit. Example (4) illustrates named-entity structure, (5) illustrates reentrancy for coreference, and (6) illustrates negation as a polarity attribute. Representative PENMAN encodings for these examples appear in Figure 3.

- (4) *Elle visited Minneapolis.*
- (5) *The boy saw himself.*
- (6) *Anna did not arrive.*

```
% Named entities
(v / visit-01
  :ARG0 (p / person :name (n / name :op1 "Elle"))
  :ARG1 (c / city :name (m / name :op1 "Minneapolis")))

% Coreference via reentrancy
(s / see-01
  :ARG0 (b / boy)
  :ARG1 b)

% Negation via polarity
(a / arrive-01
  :ARG1 (p / person :name (n / name :op1 "Anna"))
  :polarity -)
```

Figure 3 AMR conventions in PENMAN notation for (4), (5), and (6).

4.2 UMR as a Cross-Linguistic, Document-Level Extension

UMR begins with AMR as a practical sentence-level representation and extends it in two directions. It is explicitly cross-linguistic, and it includes a companion document-level representation that links meaning across sentences (van Gysel et al. 2021, 2022). A key motivation is that many interpretive

phenomena are not sentence-bounded. Narratives require temporal linking; argumentation requires tracking who is committed to which claims; and reference tracking requires cross-sentential coreference. UMR treats these dependencies as first-class objects, not afterthoughts.

At the sentence level, UMR adds systematic encodings that are useful for cross-linguistic comparison, including aspect and person/number features for referential expressions (van Gysel et al. 2022). At the document level, it provides explicit structures for coreference, temporal dependencies, and modal (commitment) relations. The first public UMR dataset and infrastructure released multilingual annotations for six languages (Arapaho, Chinese, English, Kukama, Navajo, Sanapanã) (Bonn et al. 2024). Figure 4 provides a minimal illustration.⁵ It is intentionally small, but it shows the central idea. Sentence-level graphs remain readable, while cross-sentence interpretation is captured in explicit links.

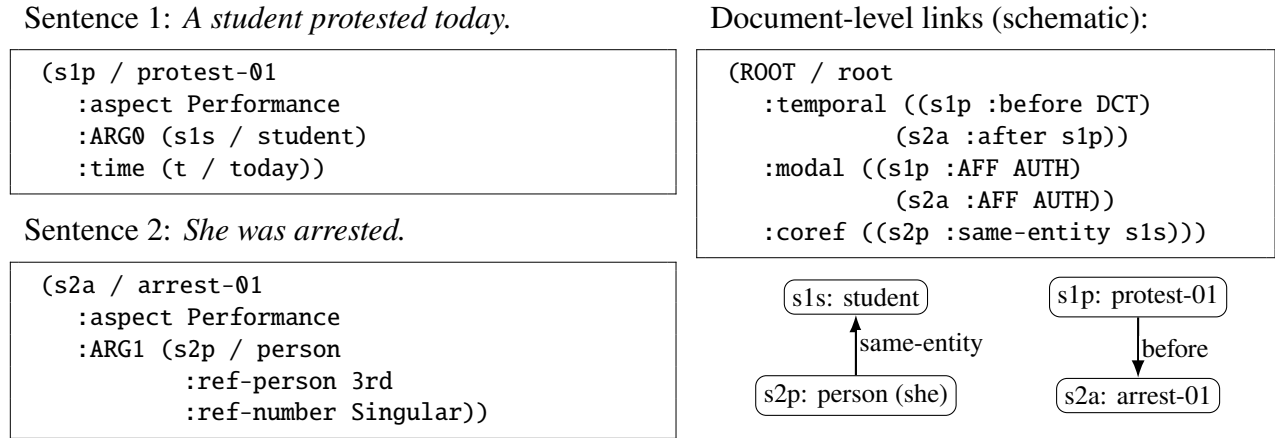


Figure 4 A compact UMR illustration with sentence-level graphs and document-level links.

4.3 Ecosystems and Multilingual Expansion

From a humanities perspective, a representation becomes methodologically useful when it comes with (i) annotation guidelines, (ii) corpora, (iii) software for reading and manipulating graphs, and (iv) baseline parsers that produce graphs good enough to support human-in-the-loop correction.

For AMR, the main English meaning bank is AMR 3.0 (LDC2020T02), released in 2020 with over 59,000 sentences across genres (Knight et al. 2020). More recently, LDC released machine translations of a subset of AMR 3.0 into Spanish, Irish Gaelic, and Dutch (LDC2024T11), reflecting continued interest in multilingual expansion and cross-lingual evaluation (Vanroy 2024).⁶ For UMR,

⁵ Document-level links include coreference, temporal ordering, and modal commitment. The figure is adapted from the original UMR documentation.

⁶ LDC refers to the Linguistic Data Consortium, a research consortium that develops, curates, and distributes linguistic data resources for research and education. It is a primary source of annotated datasets used in computational linguistics and natural language processing. LDC identifiers such as LDC2020T02 and LDC2024T11 are catalog numbers used by

the first public multilingual dataset and infrastructure were released in 2024 (Bonn et al. 2024), and the “toolchain” continues to expand through work on UMR parsing and UMR-to-text generation (Chun and Xue 2024; Markle et al. 2025a,b).

4.4 What AMR and UMR Leave Implicit (and Why That Is Methodologically Useful)

A frequent misunderstanding is that a meaning representation should encode *all* meaning. Humanities scholars will immediately see why this is neither possible nor desirable. Interpretation is not a finite list of features, and texts routinely convey meaning through what is implied, presupposed, or rhetorically performed. The question is therefore not whether AMR/UMR are complete, but whether their incompleteness is *disciplined*.

Several kinds of meaning that often matter for humanities interpretation remain underspecified in sentence-only graphs. Standard AMR does not systematically encode many discourse- and pragmatics-heavy phenomena that matter in humanistic interpretation. Some representative examples include implicatures, presuppositions, irony, genre conventions, metaphor, and many discourse relations. Even UMR, while richer at the document level, is not intended to be a complete theory of narrative or rhetoric. For example, a sarcastic utterance may have an AMR that looks indistinguishable from a sincere one, because sarcasm is a pragmatic stance rather than a change in core event structure.

The idea of a thin semantic core is a methodological decision. Rather than treating this as a failure, a humanities-centered workflow can treat it as a design principle. A thin semantic core supports two kinds of scholarly practice. It makes certain commitments explicit and checkable, including events, participants, attribution, and temporal ordering, and it leaves room for interpretive pluralism by keeping higher-level readings as linked layers rather than as forced commitments in the core. This mirrors familiar humanities practice, where scholars distinguish between what a text *says*, what it *implies*, and what it *does* rhetorically.

Layering is a practical compromise between interpretive richness and annotator agreement. The practical problem with adding more semantics is that annotation becomes harder and less reliable. This is not merely an engineering inconvenience; it is also a humanities concern, because low reliability can hide disagreement. Linked layers provide a compromise. A scholar can annotate stance, affect, or framing as separate structures attached to nodes or subgraphs, and can explicitly represent uncertainty or competing analyses. This style of layered annotation treats the graph as a scholarly object. It makes the semantics of the core explicit while keeping higher-level readings contestable and citable. It aligns with recent work on reproducibility and explainability in digital humanities and with proposals that evaluate generative AI systems as cultural technologies that require interpretive accountability (Ries et al. 2024; Kommers et al. 2026). To make the division

the Linguistic Data Consortium to identify and distribute specific datasets. They function as stable references to particular releases rather than as descriptive document titles.

between a thin semantic core and richer interpretive layers concrete, Table 1 summarizes which analytic concerns are typically supported directly by AMR and UMR and which concerns are better treated as linked layers.⁷

Table 1 A schematic comparison of semantic commitments and interpretive concerns.

Analytic concern	Typical AMR support	Typical UMR support	Likely added layer
Event structure; roles	Strong (PROPBANK roles)	Strong + aspect features	Agency framing; responsibility
Reference tracking	Within-sentence only	Cross-sentence coreference	Character/narrative roles
Temporal sequencing	Limited; often implicit	Document-level temporal links	Plot structure; pacing
Attribution/commitment	Limited conventions	Document-level modal relations	Evidentiality; stance nuance
Style and rhetoric	Not represented	Not represented	Voice, framing, genre
Affect and evaluation	Partially lexicalized	Partially lexicalized	Emotion/affect annotations

5. MEANING GRAPHS IN THE LLM ERA

Meaning graphs matter in the LLM era in two complementary ways. First, LLMs can be used to propose graphs (semantic parsing). Second, graphs can be used to guide or audit LLM outputs (generation and interpretation).

5.1 LLMs as Semantic Parsers: Promise and Limits

In principle, prompting could make semantic parsing accessible. One might imagine asking a model to output PENMAN graphs directly. However, empirical studies suggest that out-of-the-box prompting is not yet reliable enough when the goal is fully accurate semantic graphs. Ettinger et al. (2023) show that GPT-style models can often reproduce the format of AMR and recover some core predicate–argument structure, but they make major errors, and even with few-shot demonstrations they have near-zero success in producing fully correct parses. More recent evaluation likewise argues that GPT models are not suitable as standalone AMR parsers, connecting the limitation to AMR-specific structures, complex semantic roles, and (crucially for humanities work) the faithfulness of explanations (Li and Fowlie 2025).

⁷ The table is illustrative rather than exhaustive. It shows how a relatively thin semantic core can support richer humanities-oriented interpretive layers.

These findings do not imply that AMR/UMR are irrelevant for humanities. They suggest that the LLMrole should be reframed. A helpful analogy is archival work. A model can help with OCR or metadata suggestions, but scholars validate critical details. Similarly, LLMscan propose candidate graphs or partial analyses to identify the main event and its participants, while humans verify and correct. In that workflow, the graph becomes the record of analysis, which is the object others can reference and scrutinize.

5.2 Graphs as Constraints and Collaborators in LLM Prompting

There is also a complementary direction in which meaning graphs help LLMS, rather than LLMS producing graphs. In low-resource settings, UMR has been explored as a prompt ingredient for improving LLM translation. Wein (2025) reports that incorporating UMRdemonstrations into GPT-4 prompts can improve translation performance for Indigenous languages in many test cases. This is a different kind of contribution. The representation functions as structured context that constrains or guides generation.

A parallel line of work focuses on text generation from UMR graphs, a prerequisite for using graphs as interfaces for controlled multilingual generation (Markle et al. 2025a). Work on UMR parsing, including the SETUP parser, aims to make UMR graphs automatically available at scale (Markle et al. 2025b). These efforts point toward an “interpretability stack” in which graphs constrain what is generated and support inspection of what the system believed the input meant.

5.3 A Practical Workflow: Graph-Grounded Humanities Analysis

Figure 5 sketches a humanities-friendly workflow.⁸ The key point is that the graph and its layers are the scholarly objects. LLMS may participate as assistants, but the workflow is designed so that crucial claims remain auditable.

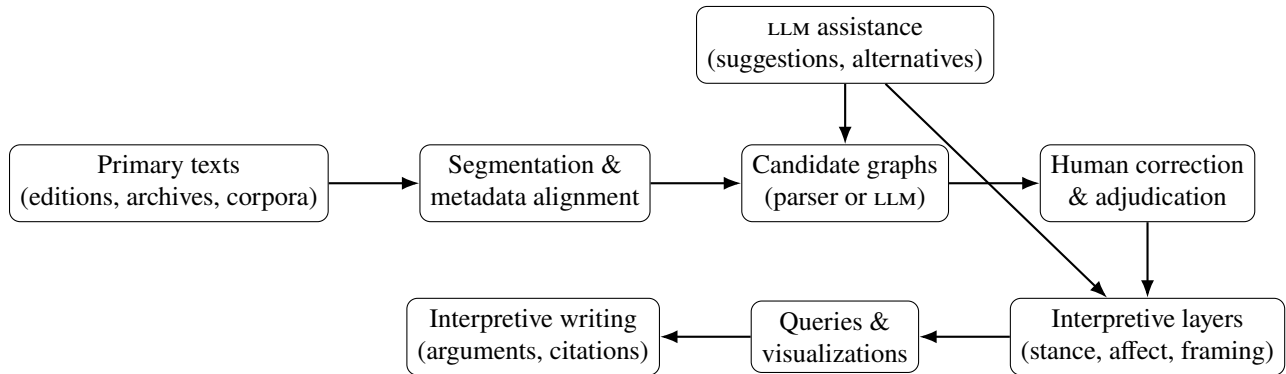


Figure 5 A graph-grounded workflow for humanities research with LLMs.

⁸ Candidate graphs and interpretive labels may be proposed by tools, but analysis and citation rely on the corrected graph and its linked interpretive layers.

6. SAMPLE ANALYSES: WHAT MEANING GRAPHS ENABLE

This section provides compact analyses. They are not meant as definitive AMR/UMR annotations. Their purpose is to illustrate the kinds of scholarly moves that become easier when semantic structure is explicit.

6.1 Agency, Responsibility, and Rhetorical Voice

Historians, rhetoricians, and critical discourse analysts often care about how texts distribute agency. Passive voice, nominalizations, and impersonal constructions can shift attention away from actors. A well-known challenge for computational analysis is that surface patterns can be misleading. A passive does not always erase agency, and an active does not always foreground it. A meaning graph can help separate two questions, namely what event-participant structure is licensed by the text and how the surface form frames that structure.

Consider the contrast in (7) and (8). A simplified shared AMR core for both examples is given in Figure 6.

(7) *The police dispersed the crowd.*

(8) *The crowd was dispersed.*

```
(d / disperse-01
  :ARG0 (p / police)
  :ARG1 (c / crowd))
```

Figure 6 *Simplified shared AMR core (PENMAN notation) for (7) and (8).*

If we treat this graph as a semantic core and add a separate “voice” layer (e.g., `voice=active` vs. `voice=passive` tied to the predicate realization), we can ask questions that are recognizably humanistic, such as which genres or institutions systematically omit `:ARG0` realizations, even when the underlying event implies an agent? When an `:ARG0` is omitted, does the surrounding discourse later supply an implicit agent through anaphora, institutional knowledge, or blame attribution? The graph provides a stable representation of participant structure that supports comparison across texts, while leaving room to treat rhetorical framing as its own analytic target. This directly supports the humanities insight that meaning and style are intertwined but not identical. A representation can hold one dimension stable while another becomes the object of analysis.

6.2 Reported Speech, Evidential Stance, and Historiographic Voice

Historiography is full of reported claims, contested evidence, and layered epistemic commitments. A narrative may attribute a claim to a witness while keeping the author uncommitted. A key humanities question is how texts manage these layers of voice. UMR’s document-level modal structure is designed to represent such dependencies explicitly.

Let us consider (9) and (10), which illustrate a reported claim and its later denial.

(9) *The chronicle claims that the governor ordered the arrests.*

(10) *But later reports deny that any order was given.*

Figure 7 sketches a possible UMR-style representation. The structure highlights two interpretive payoffs. First, it distinguishes events (ordering, arresting, denying) from sources (chronicle, reports). Second, it encodes modal commitment as a relation between sources and propositions. This makes it possible to ask which sources affirm or deny which events, and how those attributions evolve across the document.

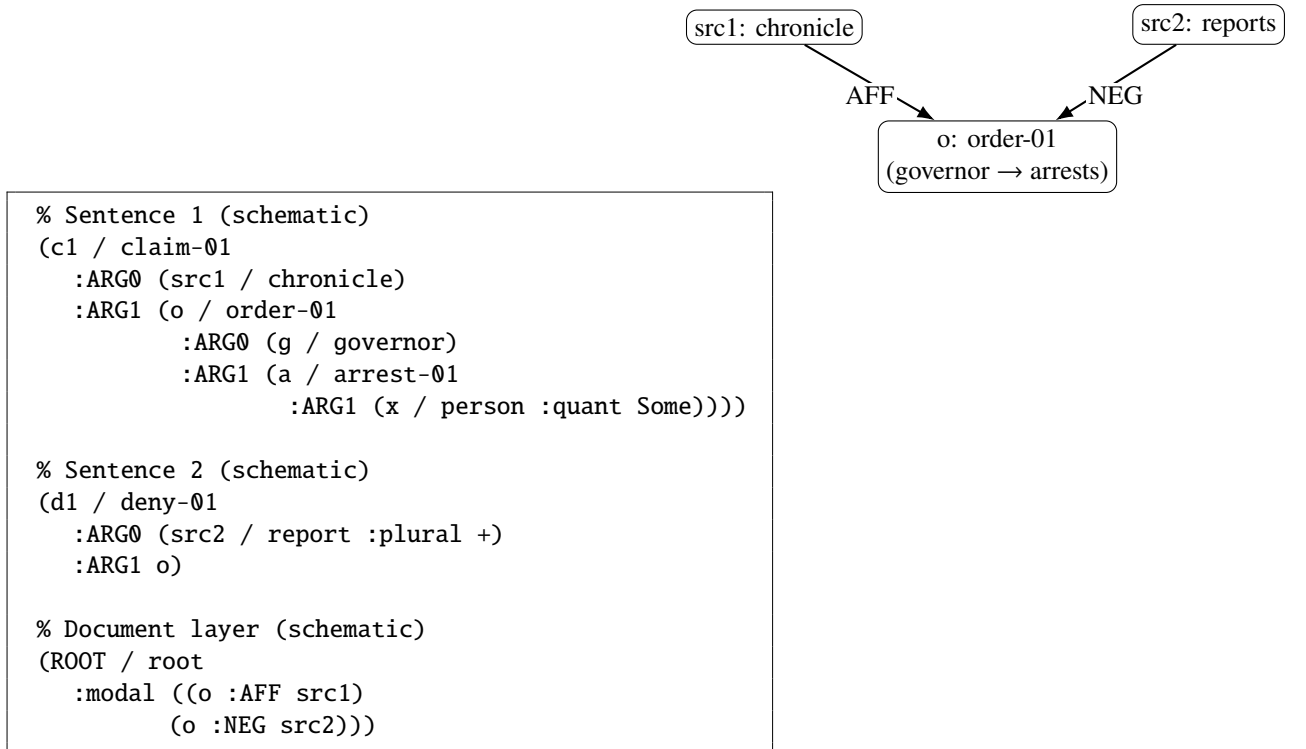


Figure 7 A schematic UMR-style sketch of competing attributions and modal commitments.

This kind of structure connects directly to humanities practices. Historians already track who

⁸ The sketch illustrates how competing attributions and modal commitments can be represented across two sentences. Labels are adapted from van Gysel et al. (2022).

claims what and with what evidential status. A graph representation makes the tracking explicit and machine-queryable. It also supports cautious LLM use. If an LLM produces a narrative summary, the researcher can compare the summary’s attributed commitments against the graph layer, identifying places where the summary collapses contested claims into asserted facts.

6.3 Narrative Temporality and Event Sequencing

Narrative analysis often depends on temporal structure, such as what happened first, what followed, what was merely intended, and what was counterfactual. Even in literary texts, ordering can be complex (flashback, foreshadowing, habitual events). Standard sentence-level representations can identify events, but they often leave discourse-level temporal relations implicit. UMR’s document layer is designed to encode such relations explicitly.

Consider (11) and (12). A simplified UMR-style sketch can encode the departure event, the regret event, and a temporal ordering (departure before regret) while also representing the causal or attitudinal link implicit in “regretted.” Such representations can support humanistic questions about narrative pacing and evaluation: Do certain genres cluster regret or judgment events after decisive actions? How often do narratives encode intentions that later fail, and how are those failures framed?

(11) *At dawn, the ship departed.*

(12) *By noon, the captain regretted the decision.*

The point here is not to claim that UMR solves narratology. The point is that even coarse temporal linking creates an explicit object that can be compared across texts and can be audited when LLMs are used for summarization or timeline construction.

6.4 Translation, Comparison, and Cross-Linguistic Semantics

Comparative humanities often depend on translation by comparing how languages express agency, temporality, or evidentiality. UMR’s cross-linguistic goal is relevant because it aims to represent shared semantic structure across typologically diverse languages while allowing language-specific detail where needed (van Gysel et al. 2021; Bonn et al. 2024).

Examples (13) and (14) offer a minimal English–Korean pair for illustration. The Korean sentence uses the resultative construction *-e iss-* to emphasize the resulting state.

(13) *The letter has been translated.*

(14) *ku phyenci-ka penyek-toy-e iss-ta.* (“The letter is in a translated state.”)

Figure 8 shows a schematic UMR-style representation that encodes the translation event and a resulting state while keeping the core participant structure stable.

```

(t / translate-01
  :aspect Performance
  :ARG1 (1 / letter)
  :result (s / state
    :aspect State
    :ARG1 1))

```

Figure 8 A schematic UMR-style representation for (13) and (14) (illustrative).

This example is not intended to adjudicate Korean aspect. Instead, it shows how a shared graph can support comparative questions. A scholar can annotate (or parse-and-correct) parallel passages, then ask where translations systematically shift aspectual viewpoint, agentivity, or attribution. This can be connected to interpretive hypotheses about genre, ideology, or historical context.

Such workflows are increasingly realistic. LDC has released machine translations of AMR 3.0 subsets into multiple languages (Vanroy 2024), and UMR datasets include low-resource and typologically diverse languages (Bonn et al. 2024). Moreover, prompt-based work suggests that UMR can be operationally useful for low-resource translation in LLM settings (Wein 2025).

6.5 Oral History and Testimony: Preserving Attribution and Evidential Chains

Oral history and testimony raise interpretive and ethical challenges that are sharpened by LLM use. Summaries can easily collapse speakers, merge episodes, or “improve” coherence in ways that distort evidence. Recent work explores LLMs for oral history understanding through tasks such as text classification and sentiment analysis, highlighting both potential and the need for careful validation (Cherukuri et al. 2025). From a humanities perspective, the core requirement is not only classification accuracy, but preservation of attribution, temporal ordering, and uncertainty.

Consider the micro-excerpt in (15) and (16). A graph-grounded approach can represent at least three distinctions that matter for historical interpretation. It can represent the arrival event, the burning event, and the difference between the narrator’s memory report and the sister’s reported speech. A schematic UMR-style structure might treat “remember” and “say” as attitudes that introduce propositions with explicit sources. The document layer can then track which propositions are affirmed by which sources, and it can represent coreference (“they” → soldiers). If an LLM later produces a summary that attributes the burning directly to the narrator, the mismatch becomes detectable. The graph encodes that the burning was reported via the sister.

(15) *I remember the soldiers arrived at night.*

(16) *My sister said they burned the village.*

This example also illustrates why meaning graphs should be treated as infrastructure rather than as

replacements for interpretation. A historian still evaluates the credibility of testimony and the social conditions under which it was recorded. The contribution of the graph is that it makes attribution and temporal structure explicit enough to be audited, queried, and compared across interviews without flattening them into an untraceable narrative.

6.6 From Close Reading to Graph Queries: Scaling without Erasing Interpretation

Graph-grounded reading does not mean abandoning close reading. It means enabling certain kinds of comparison that complement close reading. Once a corpus is represented as graphs (even partially, with human correction for critical passages), scholars can write queries that correspond to interpretive hypotheses.

For example, a project on political rhetoric might query for coercion-related events by searching for particular predicate labels in the graphs, such as `arrest-01` or `ban-01`, and then check whether the agent role is explicitly realized or systematically backgrounded. One way to do this is to compare passages where the same semantic role (for example, the agent-like role `:ARGO`) is present in the graph but is not overtly expressed in the text, as in passive or impersonal constructions. A project on historiographic voice might query for document-level modal links that separate what is reported from what the author endorses. In UMR-style notation, a document layer can record entries of the form “source SRC affirms proposition p,” which can be sketched as `:modal (p :AFF SRC)`. Here `p` stands for an event or proposition node in the meaning graph, `:AFF` indicates affirmation, and `SRC` is the reporting source. Queries of this kind can retrieve places where a chronicle, witness, or later report is committed to a claim while the author remains uncommitted. A project on translation might query for systematic shifts in aspectual viewpoint. For example, we may observe cases where one language encodes a resulting state (or another aspectual distinction) that the other language omits, even when the core event-participant structure is aligned. Such queries do not produce interpretation automatically. They produce *evidence candidates*: subsets of passages where an interpretive phenomenon is likely present. The humanities value is that scholars can then perform close reading on those subsets while maintaining an explicit record of how the subset was selected.

7. A HUMANITIES-CENTERED RESEARCH AGENDA

If AMR/UMR are to function as humanities infrastructure, two research directions become central. The first concerns *representational extensions*. This means which additional meaning distinctions are needed for interpretive work, and how they can be encoded without sacrificing annotatability. The second concerns *method and evaluation*. These concern how we validate graph-grounded analyses as humanities arguments, not only as NLP benchmarks.

7.1 Extending Beyond the Core: Stance, Affect, and Narrative Structure

Even UMR, while richer than AMR, is not a complete model of rhetoric or narrative. Humanities research often requires additional layers. Humanities interpretation often turns on meaning components that a sentence-level semantic graph does not normally encode as part of its core. These include stance and evaluation, such as approval, condemnation, irony, and distance; affect and emotion, including how emotions are attributed, oriented, or framed across narratives; and narrative roles and plot structure, where events and participants are interpreted as taking on roles such as victim, hero, or villain and as participating in plot patterns. These are interpretive constructs that are often contested, so it is methodologically safer to represent them as linked layers rather than as irreversible commitments in the semantic core.

A graph-grounded workflow suggests treating these as *linked interpretive layers* rather than forcing them into the core semantic graph. This move echoes a core humanities norm. Interpretations should remain grounded in textual evidence and open to contestation. Graph-grounded reading therefore treats interpretive layers as explicit proposals linked to evidence, rather than as hidden assumptions. The humanities contribution is not only to add new labels, but to develop explicit evidential norms for how interpretive layers relate to text and to one another, especially when LLMs are involved (Kommers et al. 2026; Hutchinson 2024).

7.2 Design Principles for Humanities-Oriented Meaning Graphs

If meaning graphs are to become humanities infrastructure, their design should be guided by humanities norms. I propose six principles, framed as methodological desiderata rather than as engineering requirements. A first principle is minimal semantic commitment with explicit justification. The core layer should encode distinctions that can be justified as broadly useful and reliably annotatable, and omissions should be explicit so that scholars know what the graph cannot support. A second principle is linkability to evidence. Graphs should be anchored to the text spans and editions they are derived from, because in humanities work “the same text” can correspond to different editions and witnesses. A third principle is layered interpretive pluralism. Higher-level interpretive categories should be represented as layers that can coexist, allowing competing analyses rather than forcing a single gold label. A fourth principle is versioning and citability. Graphs and layers should be versioned so that scholars can cite the exact structures used in an argument, analogous to citing an edition or archival item. A fifth principle is transparency about tooling. When LLMs or parsers propose annotations, the proposal should be recorded, including prompts, model identifiers, and revision history, supporting reproducible critique (Ries et al. 2024). A sixth principle is ethical community alignment. For low-resource and Indigenous languages, representational work intersects with community goals and data governance, so collaboration and respect for language communities are part of representational methodology, not an optional add-on.

These principles treat meaning graphs as interpretive infrastructure rather than as mere interme-

diate data. They align with humanities discussions that analyze LLMs as reshaping epistemic agency and scholarly responsibility and that argue for evaluation frameworks grounded in interpretation rather than only in performance metrics (Seddone 2025; Kommers et al. 2026; Ries et al. 2024).

7.3 Interoperability with Humanities Data Formats and Metadata

Many humanities corpora are not “just text.” They come with rich metadata (date, place, author, edition), structural markup (chapters, stanzas, marginalia), and links to collections. A practical question is therefore how meaning graphs can interoperate with existing humanities infrastructures. One path is to treat graphs as annotations that can be linked to stable identifiers in a digital edition or archive. In that model, a graph node (an event or entity) can be connected to a span in the text, to a page image, or to an archival record. This supports two complementary research practices. Humanities scholars can move from graph queries back to original sources for close reading, and computational workflows can use metadata to stratify analyses by genre, period, region, or editorial tradition. This interoperability requirement also clarifies why UMR’s document-level layer matters. It supports reference tracking and temporal linking across units that often correspond to humanities-relevant segments (letters in a correspondence archive, interviews in an oral history project, scenes in a play). We provide representative pairings of humanities research questions, core graph evidence, and likely interpretive layers in Table 2 as a compact guide to what a graph-grounded workflow can and cannot support directly.⁹

Table 2 Examples of humanities research questions supported by graph-grounded workflows.

Humanities question	Core graph evidence	Likely interpretive layer
How is agency distributed in institutional narratives?	Predicate–argument roles; missing/implicit agents; voice metadata	Rhetorical framing; institutional responsibility
How does a historian manage competing sources?	Modal links (who affirms/denies); reported speech structures	Source credibility; historiographic stance
What is the temporal logic of a narrative?	Event graphs; document-level temporal links	Plot structure; pacing; narrative evaluation
How do translations shift aspect or evidentiality?	Cross-lingual aligned events; aspect features (UMR)	Translation strategy; ideological framing
How do oral histories preserve uncertainty and memory?	Attitude predicates (remember, say); attribution links	Ethical interpretation; trauma-sensitive reading

⁹ “Core” refers to AMR/UMR semantic structure, while “Layer” refers to linked interpretive annotations built on top of that structure.

7.4 Evaluation as a Humanities Question

NLP evaluation typically asks whether a system reproduces gold annotations. Humanities evaluation must also ask whether an annotation supports interpretation without foreclosing debate. Several criteria are relevant. Evaluation in this setting cannot be reduced to whether a system reproduces a single gold analysis. One relevant criterion is contestability, meaning that alternative analyses can be represented and compared rather than being flattened into one “correct” answer. A second criterion is traceability, which means interpretive claims can be tied back to explicit graph structures and to text spans. A third criterion is parsimony. That is, the representation does not multiply distinctions that annotators cannot apply consistently. A final criterion is community alignment, especially for low-resource and Indigenous languages, where representational choices intersect with permissions, governance, and the goals of speaker communities.

These criteria align with broader calls for reproducibility and explainability in digital humanities (Ries et al. 2024) and with recent humanities discussions of LLMs as mediating cultural technologies (Kommers et al. 2026; Seddone 2025).

7.5 Humanities Contributions to AI Development

Finally, AMR/UMR research opens a constructive pathway for humanities contributions to AI. Humanities scholars are experts in interpretation, evidence, and cultural variation. Meaning representation projects require exactly these skills, such as deciding what distinctions matter, how to represent them, and how to document the consequences of representational choices. A humanities-centered meaning representation agenda could therefore contribute to AI in at least three ways. A humanities-centered research agenda can proceed along intertwined tracks of design, critique, and building. The design track develops representational layers that capture stance, narrative roles, and other humanities-relevant phenomena while remaining computationally analyzable. The critique track treats parsers and LLMs as sources whose failures reveal training-data skews, genre biases, and cultural blind spots that matter to humanities inquiry (Ettinger et al. 2023; Hutchinson 2024). The building track develops tools, datasets, and workflows that make graph-grounded reading possible at scale, while keeping provenance, versioning, and human oversight central.

8. LIMITATIONS, RISKS, AND ETHICAL CONSIDERATIONS

Graph-grounded reading is not a cure-all. It introduces its own limitations and ethical risks, some technical and some epistemic.

8.1 Graphs Are Interpretive Objects

A first limitation is conceptual. Meaning graphs are not neutral transcriptions of meaning; they are the result of representational choices. `PROPBANK` frames, role inventories, and annotation guidelines reflect implicit theories about what counts as a predicate, what counts as an argument, and what distinctions are salient. For humanities work, this is not a reason to reject graphs, but a reason to treat them as scholarly objects that can be criticized. A graph-grounded workflow should therefore record guidelines, decisions, and revisions, and it should represent uncertainty and disagreement where appropriate.

8.2 Tool Mediation and Uneven Performance

A second limitation concerns tool mediation. When graphs are produced automatically (by parsers or `LLMs`), performance is uneven across genres, registers, and languages. Humanities corpora often include archaic orthography, dialect, code-switching, and noisy OCR. These conditions stress `NLTools` and can introduce systematic error. Human-in-the-loop correction is therefore not optional for scholarly uses; it is part of the method. Evaluation studies of `LLMs` as `AMR` analyzers provide a cautionary baseline. Fluent outputs are not evidence of semantic reliability (Ettinger et al. 2023; Li and Fowlie 2025).

8.3 Source Criticism for Models and Datasets

A third limitation is historiographic. Models and datasets embody histories of digitization and selection. Work on `LLMs` in history argues that they should be treated as sources requiring source criticism, because they draw on training corpora that are uneven in coverage and reflect social power (Hutchinson 2024). The same point applies to meaning banks. `AMR` and `UMR` corpora reflect annotation priorities, genre selections, and institutional constraints including licensing. A responsible humanities use of meaning graphs should therefore include dataset criticism, such as what is included, what is absent, and what kinds of language are systematically underrepresented.

8.4 Ethics in Multilingual and Indigenous Language Settings

Finally, `UMR` is explicitly multilingual, including low-resource and Indigenous languages (Bonn et al. 2024). This is a strength, but it also raises ethical questions about data governance, community involvement, and the purposes of annotation. Humanities scholarship has experience with collaborative and community-engaged methods, and those methods are relevant here. In such settings, a “meaning representation” project is also a cultural and political project. It can support language documentation and education, but it can also reproduce extractive research patterns if not designed collaboratively.

These limitations point to a general conclusion. Meaning graphs are most valuable when they are treated as infrastructure for responsible interpretation rather than as automated substitutes for it. They make some commitments explicit and thus make critique possible, but they do not eliminate the need for judgment.

9. RESOURCES AND RESEARCH INFRASTRUCTURE

This section lists key resources for readers who want to experiment with graph-grounded workflows. The list is intentionally explicit. A humanities methodology cannot be reproduced if the tools and datasets remain implicit.

9.1 Datasets and Corpora

A practical starting point is AMR 3.0, distributed by the Linguistic Data Consortium as LDC2020T02 (Knight et al. 2020). For multilingual and comparative work, LDC2024T11 provides automatic translations of a subset of AMR 3.0 sentences into several languages (Vanroy 2024). Researchers and students also rely on the community-maintained AMR guidelines repository, which documents conventions and updates that are not always visible from the dataset releases. These resources matter because they turn meaning graphs from an abstract proposal into a stable object of analysis that can be reused and compared across projects.

9.2 Annotation Tools and Libraries

Tooling is part of the infrastructure. The `PENMAN` library provides widely used software for manipulating AMR graphs and converting between graph objects and `PENMAN` notation (Goodman 2020). For UMR, `UMR-WRITER` supports annotation and review in a web-based interface (Zhao et al. 2021). The UMR project also maintains a public website with documentation, guidelines, and links to released datasets and tools (Uniform Meaning Representation Project 2025).

9.3 Parsing and Generation Research

Empirical work on automation and evaluation is also part of the resource landscape. Studies evaluating GPT-style models as AMR analyzers provide cautionary baselines for humanities users, showing that fluent outputs can mask major structural errors (Ettinger et al. 2023; Li and Fowlie 2025). For UMR, work on parsing and generation is beginning to establish a technological ecosystem, including pipelined parsing approaches (Chun and Xue 2024), UMR-to-text generation (Markle et al. 2025a), and English sentence-level UMR parsing (Markle et al. 2025b). Research also explores whether UMR can improve downstream tasks such as translation from low-resource and Indigenous

languages when used within LLM prompting (Wein 2025).

9.4 Humanities-Facing Resources on AI Workflows

Within digital humanities, resources include not only tools but also methodological work that documents how scholars use generative AI and how those uses should be recorded. Armaselu (2024) argues for documenting GENAI use within reproducible workflows, and Ries et al. (2024) situates these concerns within broader debates about reproducibility and explainability. Empirical studies of adoption and critique in digital humanities also provide context for how meaning graphs might be integrated responsibly, including discussions of cautious adoption and shifting scholarly roles (Ma et al. 2024; Seddone 2025). Methodological work on qualitative coding with LLMS is especially relevant because it frames interpretation as reflexive practice rather than as mere labeling (Dunivin 2025).

10. CONCLUSION

The humanities need more than access to generative tools. They need research instruments that preserve interpretive accountability. AMR and UMR are not full theories of meaning, but they are valuable as standardized, inspectable semantic artifacts. In the LLM era, their value increases rather than decreases. Empirical work shows that LLMS are not yet reliable as standalone semantic parsers (Ettinger et al. 2023; Li and Fowlie 2025), but LLMS can contribute in human-in-the-loop settings, and meaning graphs can constrain and audit their outputs. UMR’s cross-linguistic and document-level design makes it especially promising for humanities domains where translation, discourse, and historiographic voice are central.

More broadly, meaning graphs open a space for a humanities research agenda that is simultaneously critical and constructive. It is critical because it foregrounds mediation, bias, and source awareness. It is constructive because it offers concrete artifacts, namely graphs and linked interpretive layers, that can be shared, debated, and refined. If humanities scholarship is to shape AI rather than merely consume it, developing such interpretive infrastructure is a plausible place to begin.

ORCID

Chongwon Park  <https://orcid.org/0000-0003-4640-0926>

REFERENCES

- Armaselu, Florentina. 2024. “Documenting the Use of Generative AI in Digital Humanities Workflows.” Paper presented at *Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities (DARIAH Annual Event 2024)*. <https://doi.org/10.5281/zenodo.11483527>.
- Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. “Abstract Meaning Representation for Sembanking.” In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 178–186. Sofia, Bulgaria: Association for Computational Linguistics. <https://aclanthology.org/W13-2322/>.
- Bonn, Julia, Matthew J. Buchholz, Jayeol Chun, Andrew Cowell, William Croft, Lukas Denk, Sijia Ge, Jan Hajič, Kenneth Lai, James H. Martin, Skatje Myers, Alexis Palmer, Martha Palmer, Claire Benet Post, James Pustejovsky, Kristine Stenzel, Haibo Sun, Zdeňka Urešová, Rosa Vallejos, Jens E. L. Van Gysel, Meagan Vigus, Nianwen Xue, and Jin Zhao. 2024. “Building a Broad Infrastructure for Uniform Meaning Representations.” In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2537–2547. <https://aclanthology.org/2024.lrec-main/>.
- Cherukuri, Komala Subramanyam, Pranav Abishai Moses, Aisa Sakata, Jiangping Chen, and Haihua Chen. 2025. “Large Language Models for Oral History Understanding with Text Classification and Sentiment Analysis.” arXiv preprint arXiv:2508.06729. <https://doi.org/10.48550/arXiv.2508.06729>.
- Chun, Jayeol, and Nianwen Xue. 2024. “Uniform Meaning Representation Parsing as a Pipelined Approach.” In *Proceedings of TextGraphs-17: Graph-based Methods for Natural Language Processing*, 40–52. Bangkok, Thailand: Association for Computational Linguistics. <https://aclanthology.org/2024.textgraphs-1.3>.
- Dunivin, Zackary Okun. 2025. “Scaling hermeneutics: a guide to qualitative coding with LLMs for reflexive content analysis.” *EPJ Data Science* 14:28. <https://doi.org/10.1140/epjds/s13688-025-00548-8>.
- Ettinger, Allyson, Jena Hwang, Valentina Pyatkin, Chandra Bhagavatula, and Yejin Choi. 2023. “‘You are an expert linguistic annotator’: Limits of Large Language Models as Analyzers of Abstract Meaning Representation.” In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 8250–8263. Singapore: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-emnlp.553>.
- Goodman, Michael Wayne. 2020. “Penman: An Open-Source Library and Tool for AMR Graphs.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*:

- System Demonstrations*, 312–319. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-demos.35>.
- Hutchinson, Daniel. 2024. “Mapping the Latent Past: Assessing Large Language Models as Digital Tools through Source Criticism.” *Journal of Digital History* 3(1). <https://doi.org/10.1515/JDH-2023-0018?locatt=label:JDHFULL>.
- Knight, Kevin, Bianca Badarau, Laura Baranescu, Claire Bonial, Madalina Bardocz, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, Tim O’Gorman, and Nathan Schneider. 2020. *Abstract Meaning Representation (AMR) Annotation Release 3.0*. LDC2020T02. Philadelphia: Linguistic Data Consortium. <https://doi.org/10.35111/44cy-bp51>.
- Kommers, Cody, Ruth Ahnert, Maria Antoniak, Emmanouil Benetos, Steve Benford, Mercedes Bunz, Baptiste Caramiaux, Shauna Concannon, Martin Disley, James Dobson, Yali Duan, Edgar Dueñez-Guzmán, Kian Francksen, Evelyn Gius, Jonathan Gray, Ryan Heuser, Sarah Immel, Richard So, Sang Lee, Dalaki Livingston, Hoyt Long, Meredith Martin, Georgia Meyer, Daniela Mihai, Ashley Noel-Hirst, Kristen Ostherr, Deven Parker, Yipeng Qin, Jessica Ratcliffe, Emily Robinson, Karina Rodriguez, Adam J. Sobey, Ted Underwood, Aditya Vashistha, Matthew Wilkens, Youyou Wu, Zheng Yuan, and Drew Hemment. 2026. “Computational Hermeneutics: Evaluating Generative AI as a Cultural Technology.” *Front. Artif. Intell.* 9:1753041. <https://doi.org/10.3389/frai.2026.1753041>.
- Li, Yanming, and Meaghan Fowlie. 2025. “GPT Makes a Poor AMR Parser.” *Journal for Language Technology and Computational Linguistics* 38(2): 43–76. <https://doi.org/10.21248/jlcl.38.2025.285>.
- Liu, Jiangfeng, Ziyi Wang, Jing Xie, and Lei Pei. 2024. “From ChatGPT, DALL-E 3 to Sora: How Has Generative AI Changed Digital Humanities Research and Services?” arXiv preprint arXiv:2404.18518. <https://doi.org/10.48550/arXiv.2404.18518>.
- Ma, Rongqian, Meredith Dedema, and Andrew Cox. 2024. “A Dancing Bear, a Colleague, or a Sharpened Toolbox? The Cautious Adoption of Generative AI Technologies in Digital Humanities Research.” arXiv preprint arXiv:2404.12458. <https://doi.org/10.48550/arXiv.2404.12458>.
- Markle, Emma, Reihaneh Iranmanesh, and Shira Wein. 2025a. “Generating Text from Uniform Meaning Representation.” In *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, 259–271. Mumbai, India: The Asian Federation of Natural Language Processing and the Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.ijcnlp-long.16>.
- Markle, Emma, Javier Gutierrez Bach, and Shira Wein. 2025b. “SETUP: Sentence-Level English-To-Uniform Meaning Representation Parser.” arXiv preprint arXiv:2512.07068. <https://doi.org/10.48550/arXiv.2512.07068>.

- Marklová, Anna, Jiří Milička, Leonid Ryvkin, L'udmila Lacková Bennet, and Libuše Kormaníková. 2025. "Iconicity in Large Language Models." *Digital Scholarship in the Humanities* 40(4): 1203–1224. <https://doi.org/10.1093/llc/fqaf095>.
- Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. "The Proposition Bank: An Annotated Corpus of Semantic Roles." *Computational Linguistics* 31(1): 71–106. <https://doi.org/10.1162/0891201053630264>.
- Ries, Thorsten, Karina van Dalen-Oskam, and Fabian Offert. 2024. "Reproducibility and Explainability in Digital Humanities." *International Journal of Digital Humanities* 6(1): 1–7. <https://doi.org/10.1007/s42803-023-00078-7>.
- Seddone, Guido. 2025. "LLMs and the Increasing Role of the Humanities in the Digital Age." *AI & Society*. <https://doi.org/10.1007/s00146-025-02778-w>.
- Suviranta, Rosa, and Tuomo Hiippala. 2025. "Can Multimodal Large Language Models Annotate Low- and High-Level Features of Multimodal Artefacts?" *International Journal of Humanities and Arts Computing* 19(2): 127–142. <https://doi.org/10.3366/ijhac.2025.0354>.
- Uniform Meaning Representation Project. 2025. "UMR Project Website." <https://umr4nlp.github.io/web>.
- van Gysel, Jens E. L., Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O’Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. "Designing a Uniform Meaning Representation for Natural Language Processing." *KI—Künstliche Intelligenz* 35(3): 343–360. <https://doi.org/10.1007/s13218-021-00722-w>
- van Gysel, Jens, Meagan Vigus, Jin Zhao, and Nianwen Xue. 2022. "Uniform Meaning Representation, a Cross-Lingual Annotation Framework for Document-Level Semantics (Tutorial Outline)." In *Proceedings of the Language Resources and Evaluation Conference (LREC 2022)*.
- Vanroy, Bram. 2024. "Abstract Meaning Representation 3.0 – Machine Translations". LDC2024T11. Philadelphia: Linguistic Data Consortium. <https://doi.org/10.35111/b94n-1y25>.
- Wein, Shira. 2025. "Can Uniform Meaning Representation Help GPT-4 Translate from Indigenous Languages?" In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 278–285. <https://doi.org/10.18653/v1/2025.acl-short.23>.
- Zhao, Jin, Nianwen Xue, Jens Van Gysel, and Jinho D. Choi. 2021. "UMR-Writer: A Web Application for Annotating Uniform Meaning Representations." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 160–167. <https://doi.org/10.18653/v1/2021.emnlp-demo.19>