Large Meaning Models (LMMs):
Phaneron as a Language-Agnostic Structural Semantics Substrate

Anonymous Preprint

November 13, 2025

Abstract

We introduce *Large Meaning Models* (LMMs): language systems whose core is an explicit, language-agnostic *meaning substrate*—the *Phaneron*—where distinctions and connections are primitive and all nuance (roles, n-ary relations, direction, tense/aspect, discourse links) is represented as reusable *patterns* and *graph rewrites*. We formalize an LMM as an abstract data type with (i) a versioned Phaneron meaning graph, (ii) bidirectional bridges between surface and meaning, and (iii) a hybrid objective that *compresses* discourse and *predicts* future utterances/world updates. We map core linguistic phenomena to unlabeled structural motifs, outline learning/inference via constraint-driven abduction and MDL+prediction, and propose evaluations for paraphrase invariance, cross-lingual alignment, explainable QA, and long-range discourse coherence. LMMs complement LLMs: LLMs act as string transducers; LMMs are the inspectable *meaning* core.

1 Introduction

Token-next architectures blur surface form with meaning, offer weak inspectability, wobble under paraphrase, and struggle with long-range discourse state. We propose to separate *meaning* from *surface* with a single-layer, unlabeled structural substrate where semantics live as reusable patterns discovered by a rewrite loop guided by compression and prediction.

2 Background: The Phaneron in brief

The Phaneron is an unlabeled, undirected pseudograph $G = (V, E)$ with reusable *patterns* (finite subgraphs), double-pushout (DPO) graph rewrites, and a versioned event log $\mathcal{E}$. Meaning equals structural position (ideally automorphism orbits; operationally, r-hop/WL colors). A hybrid objective $\mathcal{L} = \mathcal{L}_{\mathrm{MDL}} + \lambda\,\mathcal{L}_{\mathrm{pred}}$ scores states and candidate rewrites.

3 The LMM ADT

Definition 1 (Large Meaning Model)**.** *An LMM is a tuple* $\mathcal{M} = (\mathcal{P}, \mathcal{B}_{\mathrm{in}}, \mathcal{B}_{\mathrm{out}}, \mathcal{U}, \mathcal{L})$ *where:*

- $\mathcal{P} = (G, \mathcal{E}, \mathcal{D}, \mathcal{R}, \mathcal{L}_{\mathrm{MDL}})$ *is a Phaneron instance (graph, event log, dictionary, rewrite schemas, MDL objective);*
- $\mathcal{B}_{\mathrm{in}}$ *maps surface streams to* candidate *graph deltas (rewrite proposals with alignments);*
- $\mathcal{B}_{\mathrm{out}}$ *maps meaning slices to surface realizations in a target language;*
- $\mathcal{U}$ *selects a consistent subset of proposals (constraint-driven abduction, CDA) and applies DPO rewrites;*
- $\mathcal{L} = \mathcal{L}_{\mathrm{MDL}} + \lambda_1\mathcal{L}_{\mathrm{pred}} + \lambda_2\mathcal{L}_{\mathrm{cycle}} + \lambda_3\mathcal{L}_{\mathrm{min}} + \lambda_4\mathcal{L}_{\mathrm{align}}$ *is the global objective.*

Figure 1: **LMM pipeline.** Surface streams are mapped to *proposals* (Bridge-In); the Phaneron applies versioned rewrites; Bridge-Out realizes responses.

Figure 2: **Architecture layers.** Surface constructions map through bridge patterns into language-agnostic meaning patterns; realization reverses the mapping.

4 Pattern typology for language

We use canonical unlabeled motifs for linguistic phenomena; roles are structural positions inside relation patterns.

Events and roles. Clausal semantics as an event node with role positions (agent, patient, adjuncts).

Alternations. Active/passive/dative shift map to the same meaning motif; only bridge patterns differ.

Quantification and scope. Restrictor–scope encoded structurally; ambiguities are competing pattern placements.

Negation/modality/tense. Operator patterns attach to events; scope is explicit via structure.

Coreference and anaphora. Entity nodes persist; pronouns propose merges with antecedent candidates.

Presupposition and discourse. Triggers require prior subgraphs; accommodation introduces minimal subgraphs.

5 Discourse as versioned rewrites

A conversation or document is a sequence of versioned events e_t; each utterance proposes rewrites. The state $S_t = (G_t, \mathcal{D}_t)$ evolves via CDA selection and DPO application; replay yields an explainable discourse history.

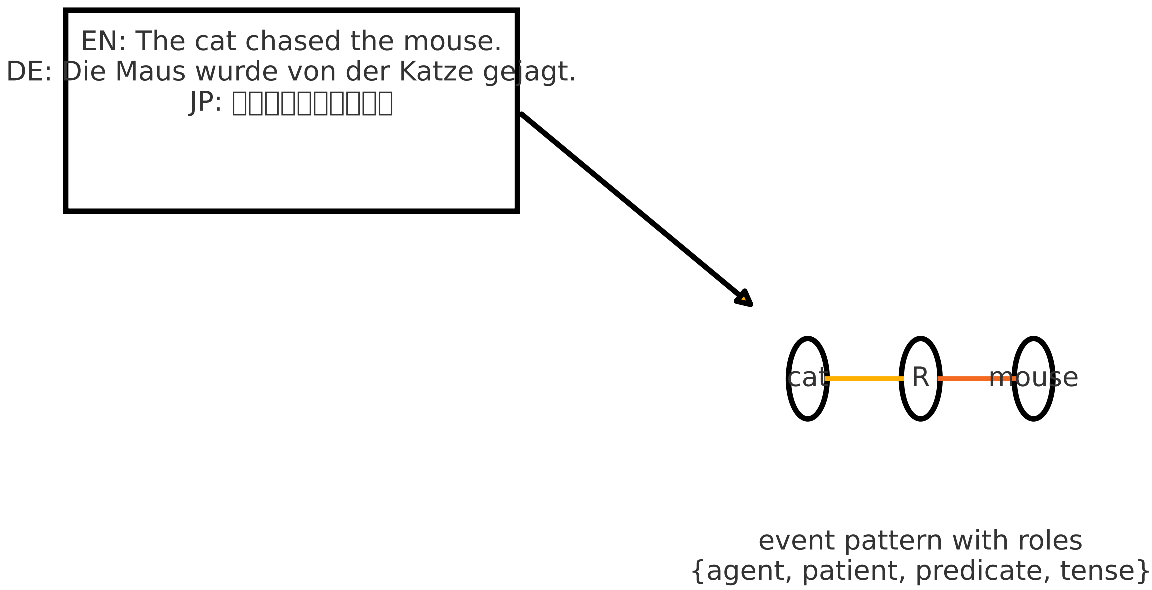

Figure 3: **Surface → meaning.** Three surface realizations (EN/DE/JP) map to a single event pattern with roles.

6 Bridges: proposal pathways and constraints

Proposal pathways. (i) Grammar-seeded parser that proposes construction-to-role alignments; (ii) LLM-assisted proposer that emits span↔role candidates; (iii) retrieval from previously accepted patterns for rapid reuse.

CDA constraints. Type and arity for roles; role-cardinality bounds; acyclicity where applicable; locality windows for new attachments; version-monotonic merges; minimality preference for explanations.

Bridge-Out. Realization chooses constructions and lexicalizations consistent with a target language and style; morphology/word order patterns fill remaining slots.

7 Learning objective (hybrid)

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MDL}}(G; \mathcal{D}, B) + \lambda_1 \mathcal{L}_{\text{pred}} + \lambda_2 \mathcal{L}_{\text{cycle}} + \lambda_3 \mathcal{L}_{\text{min}} + \lambda_4 \mathcal{L}_{\text{align}}. \tag{1}$$

8 Evaluation protocols

Paraphrase invariance. Multiple surface forms must map to isomorphic meaning patches (graph isomorphism/edit distance).

Cross-lingual alignment. Translations align to identical meaning graphs; test zero-shot transfer.

Explainable QA. Answers with Minimal Explanation Subgraphs (MES); evaluate correctness and sparsity.

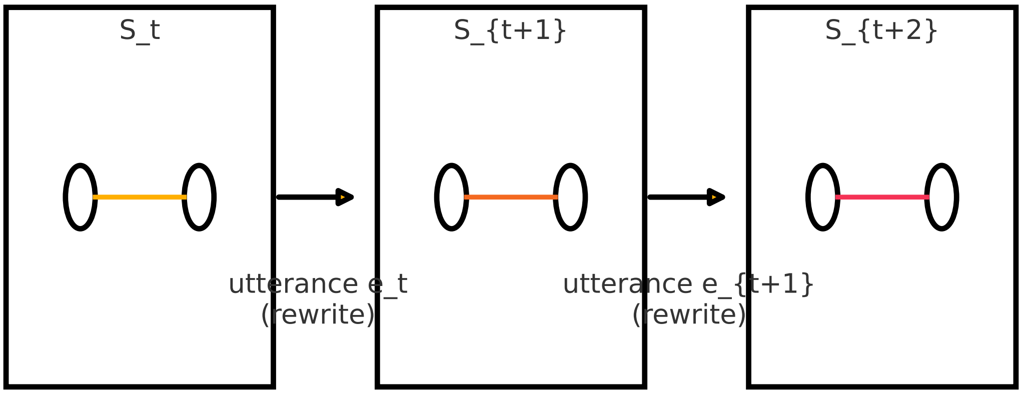

Figure 4: **Discourse timeline.** Utterances update the meaning graph by applying selected rewrites; anaphora appears as merges across versions.

Figure 5: **Inference loop.** Propose $\to$ score $\to$ select $\to$ apply.

Discourse tracking. Coreference and presupposition accommodation across long contexts.

Task	**Metric(s)**	**Structural check**
Paraphrase invariance	Graph isomorphism rate; edit distance	Canonicalization match; MES stability
Cross-lingual alignment	Zero-shot accuracy; alignment F1	Same meaning hash; role consistency
Explainable QA	EM / F1 + explanation F1	MES size/precision vs gold
Coreference	MUC/B^3/CEAF or link F1	Merge decisions; entity continuity
Discourse relations	Accuracy / F1	Correct rhetorical link patterns
Efficiency	Latency; updates/s; growth	$\lvert\mathcal{D}\rvert$ growth; replay time

Evaluation plan (summary).

9 Micro-bench: controlled paraphrase set

Dataset. 100 sentence triplets (active, passive, topicalized) per predicate, across 20 predicates; 10 languages (EN/DE/ES/FR/IT/PL/NL/SV/JA/KO) with human-checked translations.

Split. Train on half the predicates and 7 languages; test zero-shot on the rest and 3 held-out languages.

Targets. (i) Paraphrase invariance $\uparrow$ (canonical hash equality), (ii) cross-lingual isomorphism rate, (iii) MES stability under paraphrase.

Ablations. No-cycle loss; no-minimality loss; fixed pattern cap vs dynamic.

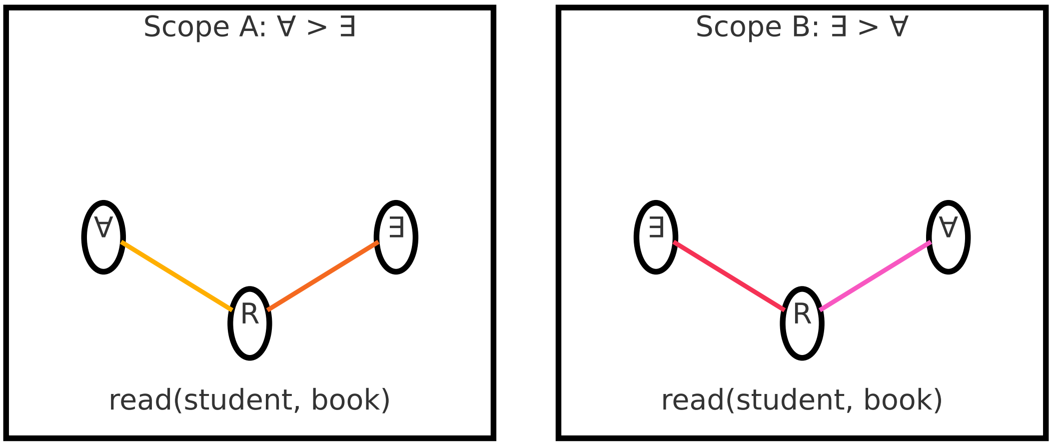

Figure 6: **Scope ambiguity.** Two competing patterns for "Every student read a book": $\forall > \exists$ vs $\exists > \forall$.

10 Case studies

Quantifier scope ambiguity. Competing structures resolved by context.

Minimal explanation subgraph (MES). "Alice dropped the plate. It shattered."

Pattern dictionary snapshot.

11 A hybrid LMM+LLM system

LLMs can serve as proposal generators and realizers; the Phaneron enforces structural constraints and provides long-term semantic memory.

12 Limitations, risks, and security

Pattern explosion, search cost, and bridge brittleness are practical risks. Explicit meaning graphs may encode sensitive facts; we recommend:

- **PII redaction**: detect and mask subgraphs linked to personal identifiers.
- **Access control**: role-based visibility of subgraph types (entities, relations).
- **Audit trails**: versioned provenance for all rewrites, with replay.
- **Safety rules**: refusal policies as subgraph pattern detectors (harmful intent).

Some pragmatic phenomena (irony, sarcasm) may require richer patterns and grounding.

13 Conclusion

LMMs make meaning first-class: explicit, stable under paraphrase, language-agnostic, and replayable. Phaneron provides the substrate; MDL+prediction provides the learning principle; versioned rewrites provide the discourse engine. LMMs complement LLMs: structure for memory/reasoning; fluency for realization.

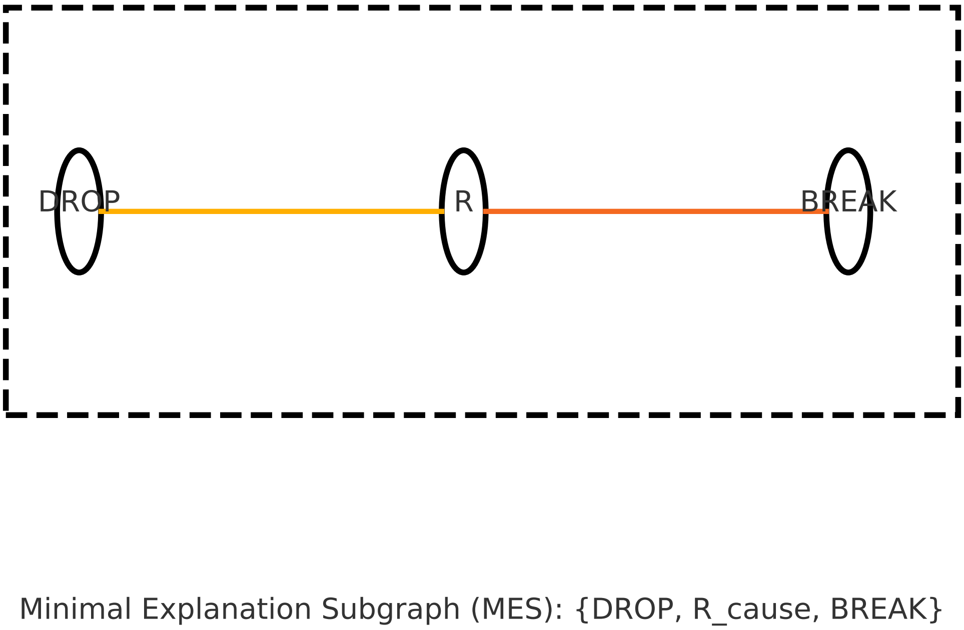

Figure 7: **Explanation view.** MES = {DROP, R_{cause}, BREAK}.

A Appendix A: Concrete MDL code

We use a two-part code for $(\mathcal{D}, B)$ and residual structure R. Let $\mathcal{D} = \{P_i\}_{i=1}^{M}$ be unlabeled patterns; U_i their usage sets (injective embeddings).

$$L(\mathcal{D}) = \sum_{i=1}^{M} \Big(L_{\mathbb{N}}(|V(P_i)|) + L_{\mathbb{N}}(|E(P_i)|) + L_{\text{iso}}(P_i) \Big), \tag{2}$$

$$L(U \mid \mathcal{D}, B) = \sum_{i=1}^{M} \Big(L_{\mathbb{N}}(|U_i|) + \sum_{u \in U_i} L_{\text{place}}(u \mid P_i, B) \Big), \tag{3}$$

$$L(R \mid \mathcal{D}, U, B) = L_{\text{edges}}(\text{residual edges} \mid B). \tag{4}$$

Here $L_{\mathbb{N}}$ is a universal integer code; L_{iso} encodes an unlabeled pattern up to isomorphism (canonical adjacency); L_{place} codes a placement via block-structured anchors B; and L_{edges} uses a Bernoulli or degree-corrected block model for uncovered edges. A rewrite is admissible if total ΔL plus predictive losses is < 0.

B Appendix B: CDA selection (pseudocode)

```
Input: state S_t=(G_t,D_t), proposals P={ G_k }, objective L(  )
1: score each  G_k  with local  L ; discard non-improving
2: sort by  L ; greedily build non-overlapping set W
3: apply DPO rewrites in W to obtain S_{t+1}; append events
4: consolidate: merge near-duplicate patterns; re-score dictionary
```

C Appendix C: Canonicalization and tests

We canonicalize small meaning patches for fast paraphrase alignment using a hash of canonical adjacency; ties are broken by lexicographically minimal rooted traversal. Paraphrase invariance

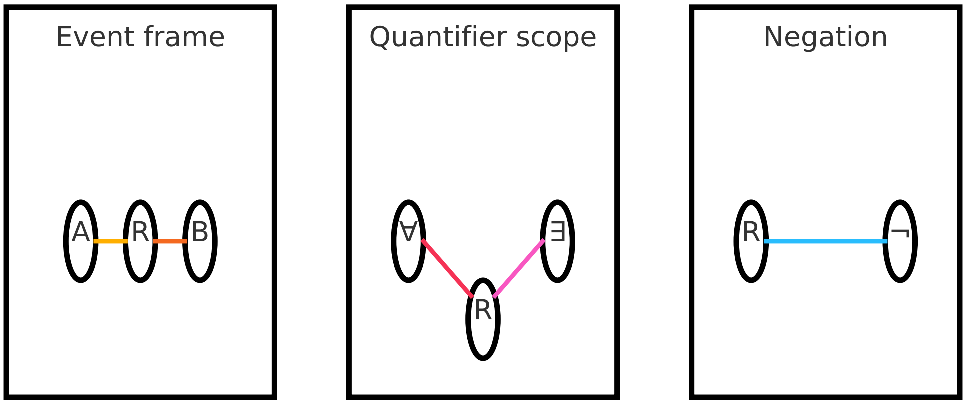

Figure 8: **Pattern dictionary motifs.** Event frame, quantifier scope, and negation attachment.

Figure 9: **Hybrid stack.** LLM front/back ends surround the explicit meaning core.

= identical canonical hashes (with audit trail).

D References

References

[1] J. Rissanen. "Modeling by shortest data description." 1978.

[2] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. 1989.

[3] B. Weisfeiler and A. Leman. "The reduction of a graph to canonical form and the algebra which appears therein." 1968.

[4] H. Ehrig et al. *Fundamentals of Algebraic Graph Transformation*. 2006.

[5] C. Fillmore et al. "FrameNet and frame semantics." 2003.

[6] L. Banarescu et al. "Abstract Meaning Representation for sembanking." 2013.

[7] N. Abzianidze et al. "UCCA: Universal Conceptual Cognitive Annotation." 2017.

[8] H. Kamp and U. Reyle. *From Discourse to Logic*. 1993.

[9] N. Asher and A. Lascarides. *Logics of Conversation*. 2003.

[10] A. Goldberg. *Constructions: A Construction Grammar Approach*. 1995.

[11] P. Lewis et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP." 2020.

[12] R. Besold et al. “Neural-Symbolic Integration.” 2017.