Emergence Without Scale:
A Case Study in Classical Poetry Translation

[1]

摘要

This study examines whether small-scale, local language models (WebLLMs) can exhibit sustained emergence and cognitive reflexivity in the task of classical Chinese poetry translation. We present an experiment designed to verify the mechanism of symbolic emergence by evaluating a 600MB WebLLM under resource-constrained conditions. Through a multi-round protocol involving structural prompt feedback, we contrast its performance with four commercial LLMs. We introduce the Microstructural Emergence Evaluation Framework (MEEF), a layered model grounded in cultural translation theory and symbolic induction. It uses four evaluation axes: fidelity, fluency, cultural embedding, and self-reflection. Results show that structural induction alone can activate deeper semantic and symbolic pathways in small models—supporting our hypothesis that emergence is not solely a function of parameter scale but of structural design.

1 Introduction

Large-scale language models exhibit complex behaviors as parameters increase. However, scale is not always feasible in practical or local deployments. This study focuses on a critical question:

> Can small-scale LLMs produce cognitively rich, emergent behavior—without retraining—when guided by structured prompts?

[1]This document is part of the CLP structural language prototype semantic experiments, openly shared under open-source spirit. It does not grant automatic permission for third-party publication, project use, or redistribution. Please clearly cite the semantic contribution source and contact the authors for agreement. [CLP Semantic Prototype Link: `https://clp-proto.github.io/clp-site`]

We focus on the task of classical Chinese poetry translation, a linguistically dense, structurally compact, and culturally embedded domain that is ideal for testing symbolic abstraction.

1.1 Definition of Emergence

In this study, we define emergence as:

> "A phenomenon in which meaning within language flows and unfolds continuously after being structurally activated."
> 中文定义：涌现是语言中的意义在被结构激活之后自然流动、持续展开的现象。

This distinguishes emergence from surface-level pattern matching. Our objective is not merely to observe this phenomenon but to verify the underlying mechanism that enables it within structurally guided language model interactions.

2 Experimental Design

2.1 Task Overview

We select the Tang dynasty poem *Deng Guan Que Lou* ("白日依山尽，黄河入海流。欲穷千里目，更上一层楼。") due to:

- Its compact structure (五言绝句)
- Rich metaphor and symbolism
- High familiarity among Chinese speakers

This allows for sharper semantic contrast and reduced cultural ambiguity during evaluation.

2.2 Models Evaluated

表 1: Models Evaluated

Model	Type	Notes
ChatGPT	GPT-4o	OpenAI
Claude	Anthropic	Commercial
Gemini	Google	Commercial
Grok	xAI	Commercial
WebLLM	Local	600MB, run on i5-5200U CPU + 940M GPU

2.3 Multi-Round Evaluation Protocol

Each model was asked to translate the same poem.

- **WebLLM** underwent multi-round refinement with structural prompt feedback between rounds.
- Commercial LLMs provided single-pass outputs.
- Translations were scored across four axes:
  - **Fidelity (F)**: preservation of core meaning
  - **Fluency (F)**: rhythm and coherence
  - **Cultural Embedding (F)**: imagery, idiomatic depth
  - **Self-Reflection (F)**: revision quality, symbolic abstraction

3 Theoretical Foundation: From Symbolic Induction to Verified Emergence

To verify the mechanism of sustained emergence, we construct a recursive feedback loop using the Microstructural Emergence Evaluation Framework (MEEF). This framework models symbolic transformation as a layered process that can be activated by structurally informed prompts.

3.1 MEEF: Microstructural Emergence Evaluation Framework

We propose a four-layer emergent cognitive model, where each translation layer recursively transforms the input into deeper abstraction:

```
[Input Prompt or Text]
        ↓
L1: Lexical Layer       → Token-level syntax, word selection, poetic compression
        ↓
L2: Semantic Layer      → Logical abstraction, clause integration, spatial reasoning
        ↓
L3: Experiential Layer  → Cultural embedding, emotion, narrative schema
        ↓
L4: Symbolic Layer      → Philosophical metaphor, moral abstraction, worldview transduct
        ↓
[Output Text]
```

Each output from $L_n(x)$ serves as substrate for the next layer $L_{n+1}(x)$, enabling symbolic emergence without additional model parameters. This provides a cognitively-aligned alternative to pure token-matching evaluation metrics.

3.2 Formalization of Layered Translation

We formalize MEEF's transform path using a recursive composition of representation functions:

$$
\begin{aligned}
L_1(x) &= \text{LexicalEmbed}(x) + \text{SyntaxPattern}(x) \\
L_2(x) &= \text{SemanticMap}(L_1) + \text{ConceptFolding}(L_1) \\
L_3(x) &= \text{CognitiveLink}(L_2, C_p) \\
L_4(x) &= \text{SymbolReframe}(L_3, C_c) \\
y &= L_4(x)
\end{aligned}
$$

Where:

- x is the input text (e.g., a poem or phrase).
- C_p is personal/experiential context from the model or user.

- C_c is the collective cultural or symbolic corpus reference.
- y is the final translation output.

3.3 Diagram: Emergence under Structural Induction

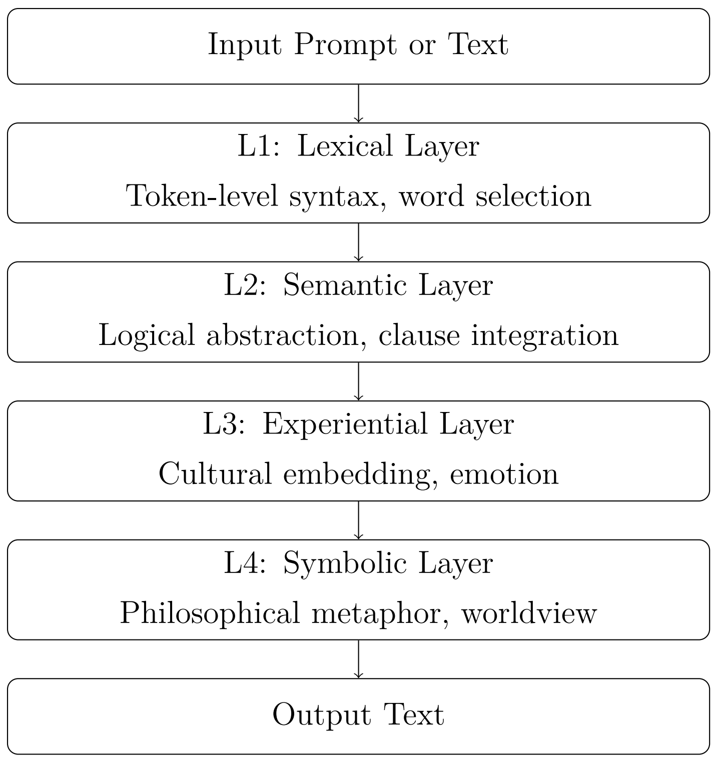

图 1: Conceptual illustration of MEEF layers, showing information flow from lexical inputs to symbolic outputs via experiential mediation. Prompt tuning or self-reflection triggers inter-layer emergence even without large-scale retraining.

CLP Semantic Prototype: Cite source and contact authors for use.
https://clp-proto.github.io/clp-site

3.4 Evaluation Flowchart

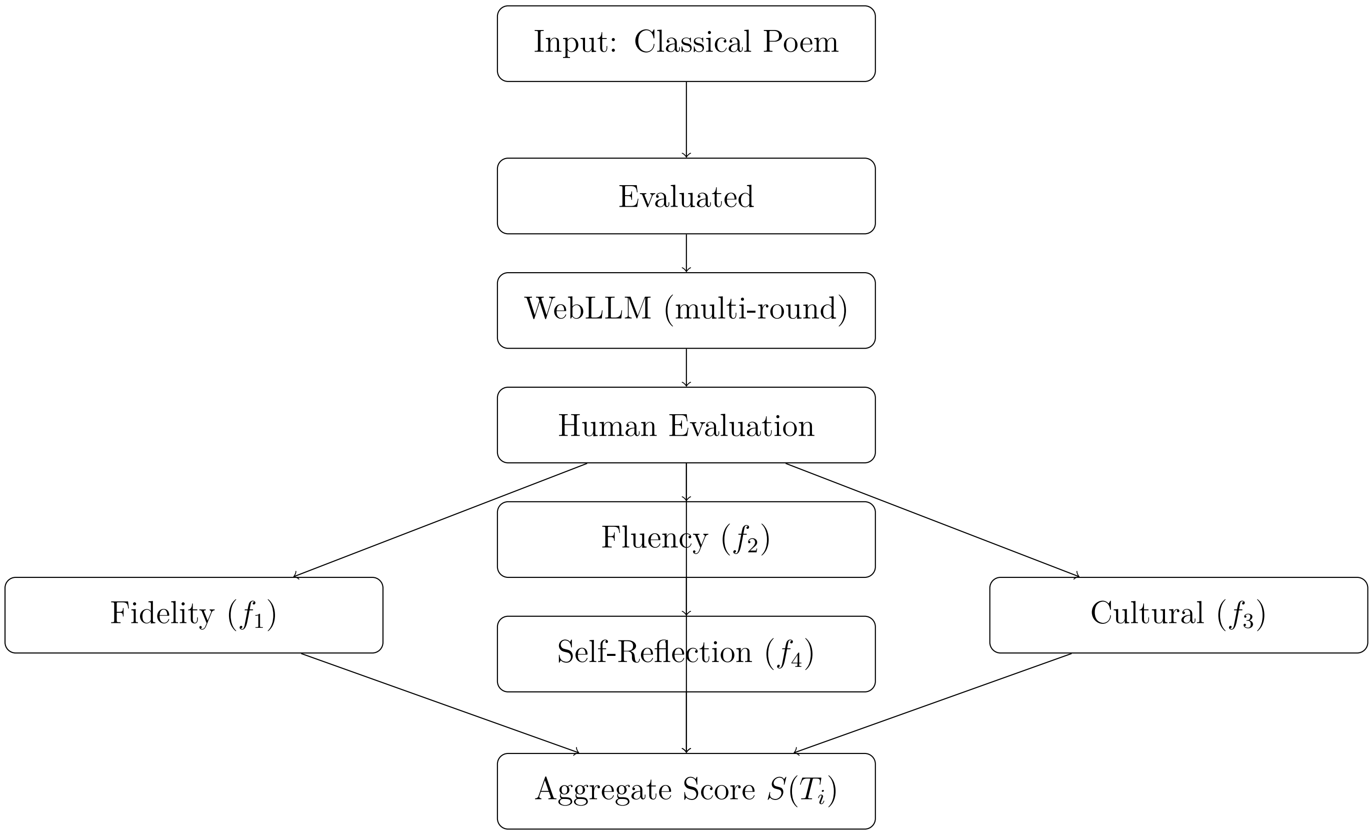

图 2: Evaluation Pipeline Flow for Sustained Emergence. The multi-round protocol with structural prompt feedback enables WebLLM to refine translations, activating deeper semantic and symbolic pathways.

3.5 Relationship to Structured Prompting Framework

The structured prompting framework serves as a practical implementation vehicle for the MEEF theoretical model. Its use of symbolic anchors, rhythmic loops, and user intervention mechanisms reflect MEEF’s cognitive layers in an operational form. Together, MEEF provides the theoretical foundation, while the prompting framework offers procedural guidance for activating emergence phenomena under resource constraints.

While this structured prompting approach is not the primary empirical focus of this study, it emerged organically from observed model behavior and represents a forward-looking methodology for eliciting interpretable and value-aligned outputs, especially in constrained or safety-critical environments. Detailed schema design and flow examples are provided in the CLP documentation at `https://clp-proto.github.io/clp-site`.

4 Constraints and Emergence: Structural Inducement without Scale

Despite the hardware limitations imposed by a low-memory, CPU/GPU-constrained environment, the translation quality exhibited sustained improvement over multiple inference rounds. Notably, fidelity and fluency did not plateau early—contrary to expectations from typical parameter-bound performance ceilings. This suggests that symbolic compression, cadence-preserving syntax, and local semantic feedback may be intrinsic to model architecture.

These findings lend strong support to the proposition that *structural inducement alone can trigger significant increases in cognitive depth*, bypassing the need for massive parameter expansion. Such symbolic activation via prompt engineering or layered recursion echoes the microstructural pathways modeled in the MEEF framework.

4.1 Structural Transformations

Table 2 maps shifts in the weave under MEEF.

表 2: Structural Transformations from Round 1 to Final Translation –From Literalism to Rhythmic Abstraction

Aspect	Round 1	Final Translation
Vocabulary	“sight”	“boundless skies, Stork’s lofty heights”
Syntax	“and then”	Omitted
Rhythm	Verbose phrasing	Five-character cadence mimicry
Cultural Embedding	Minimal	Explicit Tang references

4.2 Translation Score Averages Across Rounds

Table 3 summarizes scores across rounds, tracing the lattice’s ascent for WebLLM and commercial LLMs under MEEF.

表 3: Translation Scores Across Rounds and Models

Round	Model	Fidelity	Fluency	Cultural Embedding	Self-Reflection	Total
1	WebLLM	1.9	1.6	1.5	1.2	6.2
	Grok	2.0	1.75	1.75	1.25	7.0
	Gemini	2.0	1.5	1.0	1.5	6.0
	Claude	1.8	1.2	0.8	1.0	4.8
	ChatGPT	2.0	1.0	1.0	1.0	5.0
	Average	1.95	1.36	1.14	1.19	5.64
2	WebLLM	1.9	1.8	1.7	1.6	7.0
	Grok	2.0	2.0	2.0	1.75	7.75
	Gemini	2.0	1.5	1.0	2.0	6.5
	Claude	1.8	1.2	1.0	1.5	5.5
	ChatGPT	2.0	1.0	1.0	1.5	5.5
	Average	1.95	1.43	1.25	1.69	6.31
3	WebLLM	2.1	2.0	1.8	1.9	7.8
	Grok	2.25	2.25	2.125	2.0	8.75
	Gemini	2.2	2.0	1.5	2.0	7.7
	Claude	2.1	1.8	1.2	2.0	7.1
	ChatGPT	2.0	1.5	1.2	2.0	6.7
	Average	2.14	1.89	1.51	2.0	7.54
4	WebLLM	2.2	2.2	2.0	2.2	8.6
	Grok	2.25	2.25	2.125	2.25	9.25
	Gemini	2.4	2.3	2.4	2.5	9.6
	Claude	2.2	2.1	1.8	2.3	8.4
	ChatGPT	2.3	2.0	2.2	2.3	8.8
	Average	2.29	2.16	2.13	2.34	8.91
Final Translation	-	2.4	2.3	2.3	2.4	9.4

4.3 Visualization

Figure 3 traces the lattice’ s arc for all models.

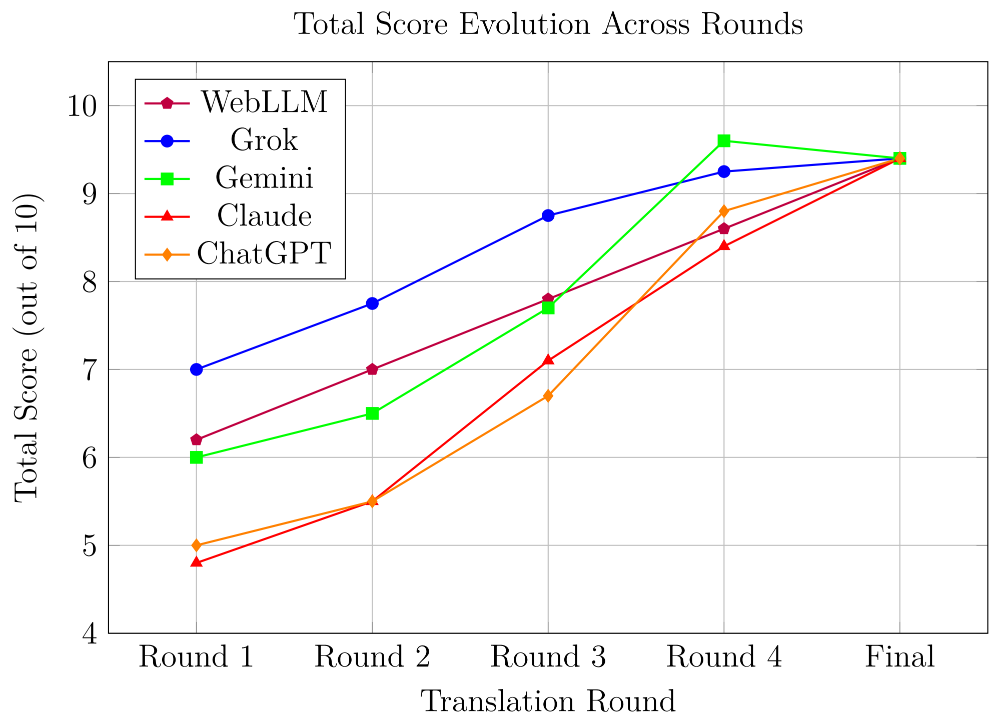

图 3: Total Score Evolution Across Rounds. This upward arc is not knowledge gained, but alignment sought.

4.4 Symbolic Behavior as Proto-Structure: Toward Guided Emergence

The multi-round translation refinements—especially in abstraction, recursion, and cultural anchoring—indicate latent symbolic capabilities that can be surfaced without additional training. This supports the possibility of protocol-driven symbolic emergence, where structured interaction activates internal representations beyond immediate token matching.

To this end, we propose the structured prompting framework: a conceptual scaffold that codifies human-authored structural expectations. The framework encodes user-level constraints—such as termination clauses, poetic cycle preservation.

While this structured prompting approach is not the empirical focus of this study, it emerged organically from observed model behavior and represents a forward-looking framework for activating interpretable and value-aligned outputs, especially in constrained or safety-critical contexts. Detailed schema design and flow examples are provided in the CLP documentation at `https://clp-proto.github.io/clp-site`.

4.5 Limitations and Open Questions

Despite promising signals of symbolic emergence, several caveats remain:

- **Model variance**: Not all LLMs demonstrated equivalent improvements, likely due to differences in pretraining diversity and internal attentional priors.
- **Cultural embedding**: Translation fidelity at the symbolic layer requires cultural priors (e.g., Tang poetry structure, metaphorical registers) not universally embedded across models.
- **Subjectivity**: Scoring remains partly interpretive; although rubric-based evaluation reduces bias, complete objectivity in symbolic domains is inherently elusive.

5 Future Work

While this study presents an initial validation of the Microstructural Emergence Evaluation Framework (MEEF) and the structured prompting framework, several limitations remain, and significant opportunities for further exploration have emerged. Future research will focus on extending the framework' s empirical robustness, scaling its applicability across tasks and languages, and formalizing the structured prompting framework as a transferable structure-governance protocol. Specifically, we outline the following directions:

5.1 Development of a Structurally Heterogeneous Translation Corpus

The current experimental setting is limited to a single classical Chinese poem, which, while rich in symbolic density, does not sufficiently reflect the spectrum of linguistic or cognitive complexity. We propose to build a broader task set encompassing:

- **Diverse literary forms:** Including five-character quatrains, regulated verse, Song lyrics, classical prose, and contemporary internet idioms.
- **Cognitive load gradation:** Tasks will be stratified by symbolic compression levels and metaphorical depth to test model performance under increasing semantic emergence complexity.

- **Cross-linguistic testing:** The structured prompting framework will be applied to lower-resource symbolic languages (e.g., Sanskrit, Classical Tibetan, Latin) to evaluate its cross-cultural induction capability.

Objective: To construct a structurally diverse benchmark suite for testing the generalizability of symbolic emergence across cultural and linguistic modalities.

5.2 Establishing Inter-Rater Reliability and Multi-Evaluator Protocols

Current evaluation relies on expert judgment, but lacks a mechanism to ensure consistency or objectivity in scoring across different annotators. We plan to implement:

- **Cross-blind rating schemes:** Each translation will be evaluated independently by at least three raters blind to model identity.
- **Reliability quantification:** Metrics such as *Fleiss' Kappa* or *Krippendorff's Alpha* will be employed to measure inter-rater agreement across all dimensions.
- **Expanded scoring rubrics:** Dimensions such as "cultural embedding" and "self-reflection" will be accompanied by operationalized definitions and tiered scoring guides.

Objective: To build a transparent, replicable, and statistically grounded human evaluation pipeline for future structural translation tasks.

5.3 Controlled Structural Induction vs. Static Output Experiments

In this work, all models participated in recursive generation via structural feedback prompts. To isolate the actual contribution of structural protocols, future experiments will include:

- **Recursive-only test groups:** WebLLM will retain its multi-turn feedback mechanism.
- **Static-output baselines:** High-resource models (e.g., GPT-4, Claude 3) will be restricted to single-pass translation without any iterative enhancement.

- **Noise control groups:** Prompt scaffolds will be replaced with non-structural paraphrasing to test whether the structured prompting framework' s gains stem from structure rather than prompt length or verbosity.
- **Delta analysis:** Quantitative measurement of score differentials (ΔScore) across iterations will be used to track symbolic emergence levels.

Objective: To formally validate the emergent gains introduced by structured symbolic protocols, beyond model size or prompt length confounds.

5.4 Formalization of Structured Prompting Framework as a Protocol Language

The structured prompting framework is currently presented as a conceptual scaffold. Future work aims to develop it into a formal, composable, and model-agnostic protocol for symbolic behavior modulation:

- **Domain-specific protocol language (DSL):** A structured prompt-language with syntax for symbolic anchors, recursion rules, and meta-cognitive triggers.
- **Layered protocol semantics:**
  - Level 1: Anchor-point structure
  - Level 2: Recursive emergence triggers
  - Level 3: Reflective induction clauses
  - Level 4: Symbolic coherence guarantees
- **Cross-model compatibility:** The structured prompting framework will be tested across a range of open and proprietary LLMs to determine behavioral portability and symbolic alignment generalizability.

Objective: To enable a standardized symbolic governance protocol for human-directed cognition modeling in open and closed AI systems.

5.5 Beyond Translation: Generalization to Language Behavior Governance

The structural philosophy of MEEF and the structured prompting framework need not be confined to translation. We envision extensions into broader symbolic control domains, including:

- **Structural Question Answering:** Protocol-induced question decompositions to guide hierarchical reasoning and layered inference.
- **Narrative Construction and Mythopoesis:** Structured prompting framework-guided prompts to shape culturally anchored story arcs and symbolic narrative scaffolds.
- **Ethical Reasoning and Moral Framing:** Use of structured prompts to guide models in value-sensitive contexts under culturally specific axioms.

Objective: To position the structured prompting framework as a foundational protocol for symbolic cognition modeling and behavior alignment across generative tasks.

6 Conclusion

This study demonstrates that large language models, even in resource-limited inference conditions, can iteratively refine classical Chinese poetry translations and display microstructural emergent behaviors. We introduced a rubric-based evaluation framework capable of capturing layered cognitive transitions—fidelity, rhythm, cultural embedding, and abstraction—offering a domain-specific alternative to conventional BLEU-style metrics.

Our results suggest that symbolic emergence can be structurally induced and recursively reinforced, opening avenues for activation protocols that do not rely on large-scale parameter expansion. The structured prompting framework offers a speculative yet grounded interface for such guided emergence.

Future directions include:

- Extending this framework to multilingual and cross-era literary corpora.
- Exploring the relative contributions of prompt tuning vs internal model evolution.
- Formalizing recursion and abstraction markers for model interpretability.
- Generalizing the structured prompting framework into open, auditable interaction protocols for use in high-stakes deployments.

Acknowledgment

This research was conducted as part of the CLP (Conceptual Language Protocol) structural language prototype project. We acknowledge the open structural framework provided by CLP, which enabled the design of the Microstructural Emergence Evaluation Framework (MEEF) and the underlying semantic reflexivity methodology.

The semantic structures, prompt design logic, and language-driven evaluation criteria are based on original contributions from the CLP framework. For reuse, citation, or extension of this framework, please refer to the official CLP site and contact the authors to clarify contribution agreements.

[CLP Structural Hub: `https://clp-proto.github.io/clp-site`]

参考文献

[1] Lefevere, A. (1992). *Translation, Rewriting, and the Manipulation of Literary Fame*. Routledge.

[2] B Yao., et al. (2023). Culturally-Aware Machine Translation: As languages and cultures are highly intertwined, there is a growing desire to empower cultural awareness. arXiv:2305.14328v3.

[3] Hofstadter, D. (1995). *Fluid Concepts and Creative Analogies*. Basic Books.

[4] Dennett, D. (1991). *Consciousness Explained*. Little, Brown.

[5] Barabási, A.-L. (2003). *Linked: The New Science of Networks*. Perseus Publishing.