Article

Not peer-reviewed version

Contextual Structural Normalization and Divergence-Based Reconstruction in Generative Models for Poetic Translation

Kalin Stoyanov *

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.1216.v1

Keywords: structural normalization; contextual knowledge state; domain-dependent divergence; analogical reasoning; poetic translation; constrained reconstruction; structural invariance; divergencebased optimization; Jensen–Shannon divergence

Article

Contextual Structural Normalization and Divergence-Based Reconstruction in Generative Models for Poetic Translation

Kalin Stoyanov

University of Chemical Technology and Metallurgy, Bulgaria; kalin.stoyanov@uctm.edu

Abstract

We introduce a divergence-based framework for structural normalization and constrained reconstruction in generative models for poetic translation. The central hypothesis is that a text admits a contextualized, language-independent structural representation capturing semantic, prosodic, rhetorical, cultural, and affective invariants independently of surface linguitic form. A normalization operator embeds each text into a domain-dependent structural manifold conditional on a contextual knowledge state $\mathcal{K}_t$. Reconstruction in a target language is formulated as divergence-minimizing projection under explicit constraint functionals. Structural preservation is quantified through domain-dependent divergence between probability measures induced by structural representations. Cross-linguistic transfer is interpreted as analogical alignment between contextualized structural states. Because structural representation depends on the contextual knowledge state, epistemic updates modify the geometry of structural comparison and may induce time-indexed optimal realizations. The proposed formulation establishes a mathematical perspective on translation as constrained structural projection in contextualized measure spaces, separating relational invariants from surface realization and enabling controllable generative reconstruction under explicit structural validation.

Keywords: structural normalization; contextual knowledge state; domain-dependent divergence; analogical reasoning; poetic translation; constrained reconstruction; structural invariance; divergence-based optimization; Jensen–Shannon divergence

---

1. Introduction

Large language models perform cross-linguistic transformations of the form

$$T : \mathcal{X}_{\ell_1} \to \mathcal{X}_{\ell_2}, \tag{1}$$

where $\mathcal{X}_\ell$ denotes the space of texts in language ℓ. These transformations operate through distributional representations learned from large corpora and are realized at the level of surface form. Although highly effective in practice, they do not explicitly separate structural invariants of a text from context-dependent realization constraints, nor do they model the dependence of interpretation on evolving external knowledge.

We introduce an alternative formulation in which cross-linguistic transfer in expressive domains is modeled as a time-indexed structural process. Let $\mathcal{K}_t$ denote the contextual knowledge state available at time t. We model translation as

$$X \xrightarrow{\mathcal{N}_{\mathcal{K}_t}} Z_{\mathcal{D},t} \xrightarrow{\mathcal{R}_{\mathcal{C},t}} X'_t, \tag{2}$$

where $Z_{\mathcal{D},t}$ denotes a domain-dependent structural representation extracted conditionally on $\mathcal{K}_t$, and $\mathcal{R}_{\mathcal{C},t}$ denotes reconstruction under a constraint set $\mathcal{C}_t$ in the target linguistic environment.

The normalization operator $\mathcal{N}_{\mathcal{K}_t}$ extracts a structural state that encodes relational and domain-relevant invariants independently of surface realization, but conditionally on available historical, biographical, and cultural information. In expressive domains such as poetry, this structural state admits a decomposition

$$Z_{\mathcal{D},t} = (S_t, P_t, R_t, C_t, A_t), \tag{3}$$

where S_t denotes semantic invariants, P_t prosodic organization, R_t rhetorical configuration, C_t cultural-historical embedding, and A_t affective structure, all interpreted relative to the contextual knowledge state.

Within the semantic component S_t, we introduce knowledge-dependent terminological anchoring. For selected lexical units τ, a canonical pivot representation

$$\pi_t(\tau) \tag{4}$$

is defined conditionally on $\mathcal{K}_t$. This pivot may correspond to a taxonomic, historically stabilized, or culturally specific identifier. The rigidity of this anchoring is modulated by a contextual freedom parameter

$$\alpha_t(\tau) \in [0, 1], \tag{5}$$

which governs the degree of permissible semantic deviation during reconstruction.

Reconstruction is formulated as a constrained optimization problem in which a domain-dependent structural divergence is minimized subject to admissibility constraints. Translation is therefore interpreted as divergence-minimizing structural reconstruction under a time-indexed contextual state. The objective is preservation of higher-order relational organization while allowing contextual adaptation as external knowledge evolves.

The contributions of this work are threefold:

- Formalization of a contextualized structural manifold $\mathcal{Z}_{\mathcal{D}}$ and a knowledge-conditioned normalization operator;
- Definition of divergence-based metrics that quantify structural invariance under evolving contextual knowledge states;
- Formulation of translation as time-indexed constrained structural projection in relational measure spaces, including knowledge-dependent semantic anchoring.

This framework provides a mathematically grounded treatment of structural equivalence in expressive translation and establishes a general principle: stable cross-linguistic transfer emerges from preservation of relational invariants conditioned on structured contextual knowledge.

2. Contextual Knowledge State and Epistemic Update

2.1. Contextual Knowledge State

Poetic translation depends not only on the surface text X, but also on externally available knowledge concerning the author, the historical period, and the surrounding cultural environment. We formalize this dependency through the notion of a contextual knowledge state.

Let t denote a time index corresponding to the currently available historical and biographical information. We define the contextual knowledge state as

$$\mathcal{K}_t = (\mathcal{H}_t, \mathcal{B}_t, \mathcal{C}_t), \tag{6}$$

where:

- $\mathcal{H}_t$ encodes historical and political context,
- $\mathcal{B}_t$ encodes biographical information about the author,
- $\mathcal{C}_t$ encodes cultural and philosophical positioning within the literary field.

 doi:10.20944/preprints202603.1216.v1

The contextual knowledge state represents the structured external information available at time t that may influence structural interpretation of the text, including the rigidity of semantic anchoring and the admissible degree of metaphorical freedom.

2.2. Conditioned Structural Normalization

Structural normalization is therefore defined conditionally:

$$Z_t = \mathcal{N}(X \mid \mathcal{K}_t), \tag{7}$$

where Z_t denotes the structural representation extracted from X under the contextual knowledge state $\mathcal{K}_t$.

Within the semantic component S_t, knowledge-dependent pivot mappings $\pi_t(\tau)$ and contextual freedom parameters $\alpha_t(\tau)$ are determined as functions of $\mathcal{K}_t$. This formalization captures the fact that semantic emphasis, affective interpretation, terminological rigidity, and cultural embedding of a poem may vary depending on available historical and biographical evidence.

2.3. Epistemic Update

We define an epistemic update as a transition

$$\mathcal{K}_t \to \mathcal{K}_{t+1}, \tag{8}$$

triggered by newly available evidence (e.g., archival discoveries, revised historical interpretations, new biographical data).

An epistemic update modifies the contextual knowledge state without modifying the original text X. Consequently, the structural representation may change:

$$\mathcal{N}(X \mid \mathcal{K}_t) \neq \mathcal{N}(X \mid \mathcal{K}_{t+1}), \tag{9}$$

and in particular

$$\pi_t(\tau) \neq \pi_{t+1}(\tau) \quad \text{or} \quad \alpha_t(\tau) \neq \alpha_{t+1}(\tau). \tag{10}$$

2.4. Time-Indexed Translation

Poetic translation is therefore time-indexed:

$$X'_t = \arg\min_Y D_{\text{poetry}}(\mathcal{N}(Y \mid \mathcal{K}_t), \mathcal{N}(X \mid \mathcal{K}_t)). \tag{11}$$

Two translations produced at different times may differ, even when the source text remains unchanged, if the contextual knowledge state has evolved. Differences may arise not only in cultural embedding or affective tone, but also in the degree of semantic anchoring enforced during reconstruction.

This formulation provides a principled mechanism for historically adaptive translation that remains structurally grounded and mathematically well-defined.

3. Core Research Questions

The contextualized structural framework introduced above raises the following research questions:

1. **Existence of contextualized structural invariants.** Does there exist a domain-dependent structural representation

$$Z_{\mathcal{D},t} \in \mathcal{Z}_{\mathcal{D}}$$

that captures relational invariants of a text conditionally on a contextual knowledge state $\mathcal{K}_t$?

2. **Stability under epistemic update.** How does the structural representation evolve under

$$\mathcal{K}_t \to \mathcal{K}_{t+1},$$

 doi:10.20944/preprints202603.1216.v1

and which components of $Z_{\mathcal{D},t}$ remain invariant across contextual revisions?

3. **Knowledge-dependent semantic anchoring.** How should canonical pivot mappings

$$\pi_t(\tau)$$

and contextual freedom parameters

$$\alpha_t(\tau)$$

be defined so that semantic precision and poetic flexibility are balanced under evolving knowledge states?

4. **Quantification of structural divergence.** How can semantic, prosodic, rhetorical, cultural, and affective invariants be quantified through domain-dependent divergence measures

$$D_{\mathcal{D}}(\cdot,\cdot \mid \mathcal{K}_t)?$$

5. **Constraint-integrated reconstruction.** How can admissibility constraints $\mathcal{C}$ be formally integrated into the reconstruction operator so that structural preservation, contextual semantic anchoring, and constraint satisfaction are jointly optimized under a time-indexed contextual state?
6. **Robustness and interpretability.** Does explicit structural normalization conditioned on contextual knowledge improve robustness, transparency, and interpretability of expressive translation systems?

4. Literature Context and Positioning

The proposed framework integrates four established research directions: structure-preserving analogy, constrained generation, information-theoretic divergence, and cross-lingual representation alignment. Its contribution lies in synthesizing these strands into a unified, divergence-based theory of contextualized structural reconstruction.

The notion that valid transfer depends on preservation of relational structure rather than surface similarity originates in structure-mapping theory [1,2] and subsequent formal accounts of analogical reasoning [3]. These works establish relational invariance as a central principle of cross-domain cognition. The present framework operationalizes this principle through a contextualized normalization operator $\mathcal{N}_{\mathcal{K}_t}$, which extracts domain-dependent structural representations conditioned on an evolving contextual knowledge state.

Constrained generation methods impose lexical, logical, or stylistic conditions during decoding [4–7]. Recent large-scale language models [8–11] achieve strong multilingual and instruction-following performance through large-scale pretraining and alignment. These systems operate primarily through distributional representations learned from corpora. In contrast, the present approach introduces an explicit structural manifold $\mathcal{Z}_{\mathcal{D}}$ and treats realization as a divergence-minimizing projection conditioned on structured external knowledge.

The mathematical formulation is grounded in information theory. Divergence measures such as Kullback–Leibler and Jensen–Shannon [12–14] provide principled metrics for comparing structural states via induced probability measures on domain-specific event spaces. In the present setting, divergence is defined conditionally with respect to a contextual knowledge state, thereby linking structural comparison to epistemic conditioning.

Cross-lingual embedding and multilingual representation learning [15–18] demonstrate that latent vector spaces can be aligned across languages. The present framework operates at a higher level of abstraction: rather than aligning embedding geometries, it defines domain-dependent structural manifolds encoding semantic, prosodic, rhetorical, cultural, and affective organization, interpreted relative to evolving contextual information.

Within the semantic component, the framework introduces knowledge-dependent canonical pivot mappings

$$\pi_t(\tau), \tag{12}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

which anchor selected lexical units to stabilized taxonomic or historically grounded identifiers. The rigidity of this anchoring is modulated by a contextual freedom parameter

$$\alpha_t(\tau) \in [0,1], \tag{13}$$

allowing the balance between semantic precision and poetic flexibility to vary across epistemic states.

Recent research has also explored regionally adapted and multilingual foundation models designed to extend base architectures to specific language ecosystems. An example is BgGPT 1.0 [19], a continually pretrained and instruction-tuned model optimized for Bulgarian language understanding and generation. The structural framework proposed here is model-agnostic and compatible with such systems. In the architectural instantiation described later, BgGPT is employed as the generative realization engine, while structural abstraction, contextual conditioning, semantic anchoring, and divergence-based validation remain external to the underlying language model.

By combining relational analogy, divergence-based comparison, contextual knowledge conditioning, semantic anchoring, and constraint-aware reconstruction, the framework formalizes poetic translation as time-indexed structural alignment within generative systems.

5. Related Work and Mathematical Positioning

The proposed framework intersects three mathematical research areas: (i) analogical reasoning as structure-preserving mapping, (ii) constrained optimization in generative systems, (iii) divergence-based alignment in domain-adaptive relational manifolds under contextual conditioning.

We position our contribution by introducing a contextualized domain-dependent structural manifold $\mathcal{Z}_{\mathcal{D}}$ and defining translation as divergence-minimizing realization within that manifold relative to an evolving contextual knowledge state $\mathcal{K}_t$.

5.1. Analogy as Structure-Preserving Mapping

Analogical reasoning has been formalized as a mapping between relational systems $\mathcal{S}_1 = (E_1, R_1)$ and $\mathcal{S}_2 = (E_2, R_2)$, where E_i denotes entities and R_i denotes typed relations.

A mapping

$$\phi : \mathcal{S}_1 \to \mathcal{S}_2 \tag{14}$$

is structure-preserving if

$$R_1(e_i, e_j) \Rightarrow R_2(\phi(e_i), \phi(e_j)). \tag{15}$$

In the present framework, structural normalization extracts a relational representation

$$Z_{\mathcal{D},t} = \mathcal{N}(X \mid \mathcal{K}_t) \tag{16}$$

that supports analogical transfer across languages. Translation is therefore interpreted as relational structure transfer conditioned on contextual knowledge and semantic anchoring, rather than as surface-form substitution.

5.2. Contextualized Structural Normalization

Let $\mathcal{X}_\ell$ denote the space of texts in language ℓ. We define a contextualized normalization operator

$$\mathcal{N} : \mathcal{X}_\ell \times \mathcal{K}_t \to \mathcal{Z}_{\mathcal{D}}, \tag{17}$$

embedding texts into a domain-dependent structural manifold relative to the contextual knowledge state.

For expressive domains such as poetry, the structural representation admits the decomposition

$$Z_{\mathcal{D},t} = (S_t, P_t, R_t, C_t, A_t), \tag{18}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

where the semantic component S_t may include pivot mappings $\pi_t(\tau)$ with contextual rigidity parameters $\alpha_t(\tau)$.

Each structural state induces a probability measure

$$P_{Z_t}^{(\mathcal{D})} \tag{19}$$

over a structured event space

$$\Omega_{\mathcal{D}} = \Omega_S \times \Omega_P \times \Omega_R \times \Omega_C \times \Omega_A. \tag{20}$$

5.3. Divergence-Based Structural Comparison

Structural comparison is defined through a contextualized divergence

$$D_{\mathcal{D}}(Z_{1,t}, Z_{2,t}) = \Delta\Big(P_{Z_{1,t}}^{(\mathcal{D})}, P_{Z_{2,t}}^{(\mathcal{D})}\Big), \tag{21}$$

where Δ may be instantiated by Jensen–Shannon divergence or Kullback–Leibler divergence. Reconstruction under constraints $\mathcal{C}$ is formulated as

$$X_t' = \arg\min_{Y \in \mathcal{X}_{\ell'}} [D_{\mathcal{D}}(\mathcal{N}(Y \mid \mathcal{K}_t), \mathcal{N}(X \mid \mathcal{K}_t)) + \lambda \mathcal{L}_{\mathcal{C}}(Y)]. \tag{22}$$

Translation is thus interpreted as divergence-minimizing realization of a contextualized structural invariant under epistemically conditioned semantic anchoring.

5.4. Relation to Constrained Generative Models

Standard constrained decoding minimizes objectives of the form

$$\mathcal{J}(Y) = \mathcal{L}_{\text{model}}(Y) + \lambda \mathcal{L}_{\text{constraint}}(Y). \tag{23}$$

The present framework introduces an intermediate structural manifold $\mathcal{Z}_{\mathcal{D}}$ and enforces invariance at the level of structural divergence $D_{\mathcal{D}}$, rather than directly at the surface token level. The generative model acts as a realization mechanism within a formally defined contextual constraint geometry.

5.5. Mathematical Novelty

The contribution of the present work lies in:

1. Introducing a contextualized structural manifold $\mathcal{Z}_{\mathcal{D}}$ conditioned on evolving knowledge states,
2. Defining divergence-based invariance relative to a contextual knowledge state $\mathcal{K}_t$,
3. Incorporating knowledge-dependent semantic pivot anchoring,
4. Formulating translation as time-indexed constrained realization in relational measure spaces,
5. Integrating structural abstraction with generative realization in a model-agnostic architecture.

This positions the framework within divergence-based optimization on domain-adaptive relational manifolds under epistemic conditioning, while enabling explicit architectural instantiation over modern large language models.

6. Structural Normalization

6.1. Conceptual Role

Structural normalization formalizes the extraction of relational invariants from a text under a given contextual knowledge state. Unlike surface-level encoding, the objective is to represent the internal structural organization of a text in a manner that is stable under cross-linguistic realization while remaining sensitive to available historical and cultural information.

Let $X \in \mathcal{X}_\ell$ denote a text in language ℓ. Let $\mathcal{K}_t$ denote the contextual knowledge state available at time t. Structural normalization is defined as a contextualized mapping

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

$$\mathcal{N} : \mathcal{X}_\ell \times \mathcal{K}_t \to \mathcal{Z}_\mathcal{D}, \tag{24}$$

where $\mathcal{Z}_\mathcal{D}$ denotes a domain-dependent structural manifold.

6.2. Time-Indexed Structural State

The structural representation of X under contextual knowledge $\mathcal{K}_t$ is

$$Z_{\mathcal{D},t} = \mathcal{N}(X \mid \mathcal{K}_t). \tag{25}$$

In expressive domains such as poetry, the structural state admits the decomposition

$$Z_{\mathcal{D},t} = (S_t, P_t, R_t, C_t, A_t), \tag{26}$$

where:

- S_t denotes semantic invariants,
- P_t denotes prosodic organization,
- R_t denotes rhetorical configuration,
- C_t denotes cultural-historical embedding,
- A_t denotes affective structure.

Within S_t, selected lexical units τ may be associated with canonical pivots

$$\pi_t(\tau), \tag{27}$$

and contextual freedom parameters

$$\alpha_t(\tau) \in [0, 1], \tag{28}$$

which regulate admissible semantic deviation during reconstruction.

6.3. Contextual Dependence

Structural normalization is explicitly conditional:

$$\mathcal{N}(X \mid \mathcal{K}_t) \neq \mathcal{N}(X \mid \mathcal{K}_{t+1}) \tag{29}$$

whenever the contextual knowledge state changes. An epistemic update

$$\mathcal{K}_t \to \mathcal{K}_{t+1} \tag{30}$$

may alter semantic anchoring, cultural embedding, or affective interpretation, even if the surface text remains unchanged.

6.4. Structural Sufficiency

The structural representation $Z_{\mathcal{D},t}$ satisfies three properties:

(a) **Cross-linguistic invariance:** it encodes relational organization independently of specific lexical realization;

(б) **Context sensitivity:** it reflects historically and biographically conditioned interpretation through $\mathcal{K}_t$;

(в) **Reconstruction sufficiency:** it contains sufficient information to generate admissible realizations in a target language under appropriate constraints.

6.5. Induced Measure Representation

Each structural state induces a probability measure

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

$$P_{Z_t}^{(\mathcal{D})} \tag{31}$$

over a structured event space

$$\Omega_{\mathcal{D}} = \Omega_S \times \Omega_P \times \Omega_R \times \Omega_C \times \Omega_A. \tag{32}$$

This induced measure provides the basis for divergence-based structural comparison and enables quantitative evaluation of translation fidelity under contextual conditioning.

7. Domain-Dependent Divergence

Let $\mathcal{D}$ denote the expressive domain under consideration. Under a contextual knowledge state $\mathcal{K}_t$, each structural representation

$$Z_{\mathcal{D},t} = \mathcal{N}(X \mid \mathcal{K}_t) \tag{33}$$

induces a probability measure

$$P_{Z_t}^{(\mathcal{D})} \tag{34}$$

over a structured event space

$$\Omega_{\mathcal{D}} = \Omega_S \times \Omega_P \times \Omega_R \times \Omega_C \times \Omega_A, \tag{35}$$

where the factors correspond to the structural components $(S_t, P_t, R_t, C_t, A_t)$.

In the semantic component S_t, selected lexical units τ may be associated with knowledge-dependent canonical pivots $\pi_t(\tau)$ and contextual freedom parameters $\alpha_t(\tau) \in [0,1]$, which regulate the admissible degree of semantic deviation during reconstruction.

This probabilistic lifting enables quantitative comparison of structural states through information-theoretic divergence.

7.1. Contextualized Structural Divergence

Structural divergence at time t is defined as

$$D_{\mathcal{D},t}(Z_{1,t}, Z_{2,t}) = \Delta\Big(P_{Z_{1,t}}^{(\mathcal{D})}, P_{Z_{2,t}}^{(\mathcal{D})}\Big), \tag{36}$$

where Δ is a divergence functional. In symmetric comparison settings, Δ is instantiated as Jensen–Shannon divergence. In reference-based settings, Kullback–Leibler divergence may be used.

Divergence is therefore explicitly conditional on the contextual knowledge state $\mathcal{K}_t$.

7.2. Semantic Anchoring and Pivot-Sensitive Divergence

For poetry, semantic fidelity includes both (i) distributional semantic preservation and (ii) preservation of knowledge-anchored referential pivots. We therefore treat $JS_{S,t}$ as a composite semantic term.

Let $\mathcal{T}_X$ denote a set of semantic units extracted from X (e.g., named entities, flora/fauna mentions, culturally loaded referents, and other pivot-candidate lexemes). For each $\tau \in \mathcal{T}_X$, $\pi_t(\tau)$ denotes the canonical pivot identifier at time t, and $\alpha_t(\tau)$ denotes the contextual freedom parameter.

We model semantic divergence as

$$JS_{S,t} = JS_{S,t}^{\text{dist}} + \mu \sum_{\tau \in \mathcal{T}_X} (1 - \alpha_t(\tau)) \cdot \mathcal{L}_{\text{pivot},t}(\tau; X, Y), \tag{37}$$

where $JS_{S,t}^{\text{dist}}$ denotes distributional semantic divergence, $\mu \geq 0$ is a weighting constant, and $\mathcal{L}_{\text{pivot},t}$ is a pivot inconsistency penalty. A canonical choice is

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

$$\mathcal{L}_{\text{pivot},t}(\tau; X, Y) = \mathbf{1}\Big[\pi_t(\tau) \not\sim \pi_t(\widehat{\tau}(Y))\Big], \tag{38}$$

where $\widehat{\tau}(Y)$ denotes the realization in Y associated with the source unit τ, and $\sim$ denotes equivalence under the pivot mapping. When $\alpha_t(\tau)$ is small, pivot preservation is enforced more strictly. When $\alpha_t(\tau)$ is large, the translation is granted greater semantic latitude.

7.3. Poetic Divergence Decomposition

For $\mathcal{D} =$ poetry, the divergence decomposes component-wise as

$$D_{\text{poetry},t} = \lambda_1 JS_{S,t} + \lambda_2 JS_{P,t} + \lambda_3 JS_{R,t} + \lambda_4 JS_{C,t} + \lambda_5 JS_{A,t}, \tag{39}$$

where $JS_{\bullet,t}$ denotes Jensen–Shannon divergence computed on the marginal distributions induced by each structural component, and $\lambda_i \geq 0$ are domain-dependent weighting coefficients.

This formulation makes explicit that semantic fidelity, prosodic alignment, rhetorical configuration, cultural embedding, and affective structure contribute additively to total structural deviation.

7.4. Time-Indexed Reconstruction Objective

Given a source text X and its contextualized structural representation $Z_{\mathcal{D},t}$, reconstruction in a target language ℓ' under constraint set $\mathcal{C}$ is formulated as

$$X'_t = \arg\min_{Y \in \mathcal{X}_{\ell'}} [D_{\mathcal{D},t}(\mathcal{N}(Y \mid \mathcal{K}_t), \mathcal{N}(X \mid \mathcal{K}_t)) + \lambda \mathcal{L}_{\mathcal{C}}(Y)], \tag{40}$$

where $\mathcal{L}_{\mathcal{C}}$ is a constraint violation functional and $\lambda > 0$ controls the trade-off between structural invariance and constraint satisfaction.

Translation is thus formalized as a divergence-minimizing projection in a contextualized structural manifold. As the contextual knowledge state evolves, both structural representation and divergence geometry may change, yielding time-indexed realizations X'_t.

8. Structural Reconstruction as Constrained Optimization

Structural normalization embeds a text X into a domain-dependent manifold $\mathcal{Z}_{\mathcal{D}}$ through the operator $\mathcal{N}(\cdot \mid \mathcal{K}_t)$, where $\mathcal{K}_t$ denotes the contextual knowledge state at time t.

Let

$$Z_{\mathcal{D},t} = \mathcal{N}(X \mid \mathcal{K}_t). \tag{41}$$

Reconstruction consists in generating a new text X'_t whose contextualized structural representation remains close to $Z_{\mathcal{D},t}$ while satisfying constraints imposed by the target environment.

Let $\mathcal{C}$ denote a constraint set encoding admissibility conditions in the target context. In the poetic domain, these constraints may include:

- metrical or rhythmic alignment,
- stylistic coherence with a literary epoch,
- preservation of affective tone,
- linguistic well-formedness,
- pivot-consistent realization of knowledge-anchored referents.

The last item captures the fact that some semantic units are anchored by contextual knowledge through canonical pivots $\pi_t(\tau)$, with admissible freedom regulated by $\alpha_t(\tau)$.

Reconstruction at time t is defined as the constrained optimization problem

$$X'_t = \arg\min_{Y \in \mathcal{X}_{\ell'}} [D_{\mathcal{D},t}(\mathcal{N}(Y \mid \mathcal{K}_t), \mathcal{N}(X \mid \mathcal{K}_t)) + \lambda \mathcal{L}_{\mathcal{C}}(Y)], \tag{42}$$

where:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

- $D_{\mathcal{D},t}$ measures contextualized structural divergence,
- $\mathcal{L}_{\mathcal{C}}$ quantifies constraint violations,
- $\lambda > 0$ controls the trade-off between structural fidelity and constraint satisfaction.

8.1. Hard and Soft Constraints

Constraints in $\mathcal{C}$ are partitioned into:

- **Hard constraints**, which define the admissible set

$$\mathcal{X}_{\ell'}^{\mathrm{adm}} = \{Y \in \mathcal{X}_{\ell'} \mid Y \models \mathcal{C}_{\mathrm{hard}}\}, \tag{43}$$

- **Soft constraints**, which contribute continuously to the penalty functional $\mathcal{L}_{\mathcal{C}}$.

Hard constraints restrict the feasible region, while soft constraints guide optimization within it.

In poetic translation, pivot constraints are typically treated as soft when $\alpha_t(\tau)$ is high and as hard when $\alpha_t(\tau)$ is low. This makes the degree of semantic rigidity an explicit function of the contextual knowledge state.

8.2. Stability and Existence

If $\mathcal{X}_{\ell'}$ is finite or compact under an appropriate topology, and if $D_{\mathcal{D},t}$ and $\mathcal{L}_{\mathcal{C}}$ are lower semi-continuous, then a minimizer exists.

Reconstruction is therefore a well-posed projection in a contextualized structural manifold.

8.3. Interpretation

Reconstruction is formulated as divergence-minimizing projection rather than token-level substitution. The contextual knowledge state $\mathcal{K}_t$ defines the geometry of structural comparison, and its evolution induces corresponding evolution of optimal realizations X_t'.

The generative model operates as a realization mechanism that explores the admissible region $\mathcal{X}_{\ell'}^{\mathrm{adm}}$, while structural divergence governs fidelity at the invariant level.

9. Domain Specialization: Poetic Structural Reconstruction

The general reconstruction framework introduced above admits domain-specific specialization. In the present work, we focus on the expressive domain $\mathcal{D} = \text{poetry}$.

9.1. Contextualized Poetic Structural State

Under a contextual knowledge state $\mathcal{K}_t$, the structural representation of a poetic text is

$$Z_{\mathrm{poetry},t} = (S_t, P_t, R_t, C_t, A_t), \tag{44}$$

where:

- S_t encodes semantic invariants,
- P_t encodes prosodic structure,
- R_t encodes rhetorical configuration,
- C_t encodes cultural–historical embedding,
- A_t encodes affective structure.

All components are conditional on $\mathcal{K}_t$, which aggregates biographical, historical, and corpus-derived contextual information.

Within the semantic component S_t, selected lexical units τ may be associated with knowledge-dependent canonical pivots

$$\pi_t(\tau), \tag{45}$$

and contextual freedom parameters

 doi:10.20944/preprints202603.1216.v1

$$\alpha_t(\tau) \in [0,1]. \tag{46}$$

The pivot $\pi_t(\tau)$ represents a stabilized referential anchor (e.g., taxonomic, historical, or culturally fixed), while $\alpha_t(\tau)$ regulates the admissible degree of semantic reinterpretation in translation. Both are functions of the contextual knowledge state:

$$\pi_t(\tau) = \Pi(\tau \mid \mathcal{K}_t), \quad \alpha_t(\tau) = A(\tau \mid \mathcal{K}_t). \tag{47}$$

Thus, semantic rigidity or flexibility is not fixed a priori, but depends on the historically and biographically informed interpretation of the poem.

9.2. Poetic Reconstruction Objective

Reconstruction in language ℓ' at time t is defined as

$$X'_t = \arg\min_{Y \in \mathcal{X}_{\ell'}} \left[D_{\text{poetry},t}(\mathcal{N}(Y \mid \mathcal{K}_t), \mathcal{N}(X \mid \mathcal{K}_t)) + \lambda \mathcal{L}_{\mathcal{C}}(Y) \right]. \tag{48}$$

Here $D_{\text{poetry},t}$ measures contextualized structural divergence, including pivot-sensitive semantic divergence, and $\mathcal{L}_{\mathcal{C}}$ encodes poetic constraints such as metrical coherence, stylistic consistency, affective alignment, and pivot-consistent realization of anchored referents.

The contextual freedom parameters $\alpha_t(\tau)$ directly influence the divergence landscape. Low values enforce strict pivot preservation, while higher values permit metaphorical or stylistic latitude. This makes semantic flexibility an explicit function of contextual knowledge.

9.3. Cultural Manifold Projection

Cultural alignment is formulated as projection onto a target poetic manifold:

$$C'_t = \arg\min_{C'' \in \mathcal{M}^{\ell'}_{\text{poetry}}} JS(C_t(X) \,\|\, C''), \tag{49}$$

where $\mathcal{M}^{\ell'}_{\text{poetry}}$ denotes the literary–historical manifold of the target language.

The cultural embedding $C_t(X)$ is itself conditioned on $\mathcal{K}_t$, reflecting the author's position within a historically evolving literary field. Consequently, cultural projection aligns not only epochal style, but also biographically informed authorial trajectory.

9.4. Context Evolution

As the contextual knowledge state evolves,

$$\mathcal{K}_t \to \mathcal{K}_{t+1}, \tag{50}$$

the induced structural state $Z_{\text{poetry},t}$, the semantic anchoring functions $\pi_t(\tau)$, the freedom parameters $\alpha_t(\tau)$, and the divergence geometry $D_{\text{poetry},t}$ may change.

Consequently, optimal realizations X'_t and X'_{t+1} may differ while preserving structural fidelity relative to their respective knowledge states.

Translation is therefore inherently time-indexed: it remains structurally faithful while adapting to epistemically updated interpretations of the author's historical, cultural, and biographical positioning.

10. Analogical Interpretation of Structural Normalization

The structural normalization framework admits an interpretation in terms of analogical reasoning. Translation is not modeled as direct symbolic transformation, but as structure-preserving transfer between contextualized domain manifolds.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

10.1. Analogy as Relational Preservation

Let X denote a poetic text situated in domain $\mathcal{D}$ = poetry, under contextual knowledge state $\mathcal{K}_t$. Structural normalization yields

$$Z_{\text{poetry},t} = \mathcal{N}(X \mid \mathcal{K}_t). \tag{51}$$

Reconstruction in language ℓ' produces

$$X'_t = \arg\min_{Y \in \mathcal{X}_{\ell'}} D_{\text{poetry},t}\big(\mathcal{N}(Y \mid \mathcal{K}_t), Z_{\text{poetry},t}\big) + \lambda \mathcal{L}_{\mathcal{C}}(Y). \tag{52}$$

Analogical transfer is therefore realized as divergence-minimizing projection of a contextualized structural invariant across linguistic manifolds.

Within the semantic component, knowledge-dependent pivots $\pi_t(\tau)$ define canonical referential anchors, and contextual freedom parameters $\alpha_t(\tau)$ regulate admissible semantic flexibility. Analogical preservation therefore includes both relational structure and pivot-consistent anchoring relative to $\mathcal{K}_t$.

Relational organization is preserved, while surface realization adapts to the admissible region of the target language.

10.2. Poetic Analogy as Cultural-Manifold Alignment

In poetry, analogy operates between culturally embedded manifolds. The structural state

$$Z_{\text{poetry},t} = (S_t, P_t, R_t, C_t, A_t) \tag{53}$$

contains a cultural embedding $C_t \in \mathcal{M}_{\text{poetry}}$.

Transfer across languages becomes alignment between cultural manifolds:

$$C'_t = \arg\min_{C'' \in \mathcal{M}^{\ell'}_{\text{poetry}}} JS\big(C_t(X) \,\|\, C''\big). \tag{54}$$

This projection preserves generational positioning, literary-historical coherence, and affective configuration through structural rather than lexical correspondence.

Semantic anchoring interacts with cultural alignment: if contextual knowledge indicates historically fixed referents, the freedom parameters $\alpha_t(\tau)$ decrease, tightening pivot preservation during analogical transfer.

10.3. Epistemic Dependence of Analogy

Because the structural state depends on the contextual knowledge state,

$$Z_{\text{poetry},t} = \mathcal{N}(X \mid \mathcal{K}_t), \tag{55}$$

analogical alignment is inherently time-indexed.

When

$$\mathcal{K}_t \to \mathcal{K}_{t+1}, \tag{56}$$

the induced structural manifold, pivot mappings $\pi_t(\tau)$, freedom parameters $\alpha_t(\tau)$, and divergence geometry may change. Consequently,

$$X'_t \neq X'_{t+1}, \tag{57}$$

while each remains structurally optimal relative to its corresponding knowledge state.

10.4. Conceptual Implications

The analogical interpretation clarifies the structure of the framework:

 doi:10.20944/preprints202603.1216.v1

- Relational abstraction precedes realization;
- Cultural manifolds encode admissible poetic configurations;
- Pivot mappings encode knowledge-anchored referential stability;
- Divergence minimization governs invariant preservation;
- The generative model performs constrained realization within an analogically aligned manifold;
- Epistemic updates modify structural geometry and may induce new optimal realizations.

Poetic translation thus instantiates a unified principle of contextualized structural transfer across heterogeneous linguistic environments.

11. Agent Architecture and Realization over BgGPT

The structural framework introduced above admits an explicit agent-based realization. The central design principle is separation between

(i) structural reasoning in the contextualized manifold $\mathcal{Z}_{\mathcal{D}}$, and (ii) generative realization in surface space $\mathcal{X}_{\ell'}$.

11.1. Layered Architecture

Under contextual knowledge state $\mathcal{K}_t$, the transformation

$$X \xrightarrow{\mathcal{N}(\cdot|\mathcal{K}_t)} Z_{\mathcal{D},t} \xrightarrow{\text{realization}} X'_t \tag{58}$$

is implemented through four coordinated modules:

1. **Structural Extractor** $\mathcal{E}$,
2. **Constraint Mapper** $\mathcal{M}$,
3. **Generative Realizer** $\mathcal{G}$,
4. **Constraint Validator** $\mathcal{V}$.

11.2. Structural Extraction

The extractor computes

$$Z_{\mathcal{D},t} = \mathcal{E}(X \mid \mathcal{K}_t) = \mathcal{N}(X \mid \mathcal{K}_t), \tag{59}$$

producing a contextualized structural state $Z_{\mathcal{D},t} \in \mathcal{Z}_{\mathcal{D}}$.

This stage includes:

- extraction of relational structure,
- estimation of cultural embedding C_t,
- computation of pivot mappings $\pi_t(\tau)$,
- estimation of contextual freedom parameters $\alpha_t(\tau)$.

11.3. Constraint Mapping

Given target language ℓ' and domain metadata, the constraint mapper constructs

$$\mathcal{C}_{\ell',t} = \mathcal{M}(Z_{\mathcal{D},t}, \ell'). \tag{60}$$

In the poetic domain, $\mathcal{C}_{\ell',t}$ encodes:

- metrical and rhythmic alignment,
- projection onto the target cultural manifold,
- preservation of affective configuration,
- pivot-consistent realization weighted by $\alpha_t(\tau)$,
- linguistic well-formedness.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

11.4. Generative Realization via BgGPT

The generative module is instantiated using BgGPT, treated as a realization operator

$$Y_t = \mathcal{G}(Z_{\mathcal{D},t}, \mathcal{C}_{\ell',t}). \tag{61}$$

Generation is conditioned through structured prompting that explicitly encodes:

- the contextualized structural representation $Z_{\mathcal{D},t}$,
- pivot anchoring instructions,
- contextual freedom parameters,
- the constraint set $\mathcal{C}_{\ell',t}$,
- a validation-compatible output schema.

BgGPT performs realization in $\mathcal{X}_{\ell'}$, while structural correctness is enforced externally. The theoretical framework remains model-agnostic; BgGPT serves as an implementation instance.

11.5. Structural Validation

Candidate outputs are evaluated through

$$\mathcal{J}_t(Y) = D_{\mathcal{D},t}(\mathcal{N}(Y \mid \mathcal{K}_t), Z_{\mathcal{D},t}) + \lambda \mathcal{L}_{\mathcal{C}_{\ell',t}}(Y). \tag{62}$$

Diagnostics include:

- semantic deviation,
- pivot inconsistency,
- prosodic inconsistency,
- cultural misalignment,
- affective drift.

11.6. Iterative Refinement

The agent operates in an iterative loop:

$$X \xrightarrow{\mathcal{E}} Z_{\mathcal{D},t} \xrightarrow{\mathcal{M}} (Z_{\mathcal{D},t}, \mathcal{C}_{\ell',t}) \xrightarrow{\mathcal{G}} Y_t \xrightarrow{\mathcal{V}} X'_t. \tag{63}$$

If $\mathcal{J}_t(Y_t)$ exceeds a tolerance threshold ϵ, validator diagnostics are injected into the next generation cycle.

Epistemic updates $\mathcal{K}_t \to \mathcal{K}_{t+1}$ trigger recomputation of structural states, pivot mappings, and divergence geometry, yielding updated realizations.

11.7. Interpretation

The architecture operationalizes the theoretical model:

- Structural invariants are extracted conditionally on $\mathcal{K}_t$;
- Pivot anchoring is explicitly modeled;
- Constraints define admissible regions in $\mathcal{X}_{\ell'}$;
- Divergence governs invariant preservation;
- BgGPT performs constrained realization;
- Epistemic updates induce structurally consistent but potentially different translations.

This separation between abstraction and realization enables controllable and interpretable structural transfer.

12. Conclusions

This work introduces a divergence-based framework for structural normalization and constrained reconstruction in generative models for poetic translation.

 doi:10.20944/preprints202603.1216.v1

Texts are embedded into a contextualized, domain-dependent structural manifold $\mathcal{Z}_{\mathcal{D}}$ through a normalization operator $\mathcal{N}(\cdot \mid \mathcal{K}_t)$, where $\mathcal{K}_t$ denotes the contextual knowledge state. Reconstruction is formulated as divergence-minimizing projection under explicit constraint sets.

The central contribution is the formal separation between relational invariants and surface realization. Semantic, prosodic, rhetorical, cultural, and affective components are encoded structurally and compared through domain-dependent divergence measures.

A key extension introduced in this work is the explicit modeling of knowledge-dependent semantic anchoring. Canonical pivot mappings $\pi_t(\tau)$ and contextual freedom parameters $\alpha_t(\tau)$ make the degree of referential rigidity an explicit function of contextual knowledge.

Biographical, historical, and corpus-derived information shape the structural state, and epistemic updates $\mathcal{K}_t \rightarrow \mathcal{K}_{t+1}$ modify the geometry of structural comparison. As a consequence, optimal translations become time-indexed realizations X'_t, reflecting the current state of contextual knowledge.

Transfer is governed by preservation of relational invariants under contextual conditioning. The generative model acts as a realization mechanism, while structural divergence and constraint validation govern fidelity and admissibility.

The framework suggests a broader principle: stable cross-linguistic transfer in expressive domains can be formalized as divergence-minimizing alignment in contextualized structural manifolds with knowledge-sensitive semantic anchoring.

Future work may explore:

- learning structural manifolds directly from corpora,
- adaptive estimation of divergence weights,
- formal probabilistic modeling of contextual knowledge states,
- empirical evaluation of robustness and controllability under explicit structural validation.

13. Normative Translation as Jurisdiction-Conditioned Structural Reconstruction

13.1. Problem Setting

We consider the drafting and multilingual realization of normative texts in a European Union setting. A legislator (e.g., a Member of the European Parliament) drafts a proposal X in a source language ℓ_s, informed by (i) the national legal tradition of the drafting jurisdiction and (ii) the supranational legal order of the European Union. The operational requirement is to produce target-language realizations X'_{ℓ_i} for multiple Member States i, where each realization must be admissible with respect to the target national legal system and coherent with the evolving EU acquis.

A key observation is that normative translation is not solely a function of the source document version. It is a function of the broader legal ecosystem: a translation may require revision even when the source text is unchanged, because related legal instruments, definitions, references, or interpretive practices have changed. Consequently, normative translation is inherently time-indexed and version-controlled.

13.2. Core Objects and Notation

Let $\mathcal{X}_\ell$ denote the space of texts in language ℓ. Let ℓ_s be the source language and ℓ_i the target language corresponding to jurisdiction i. We distinguish the source normative proposal $X \in \mathcal{X}_{\ell_s}$ from its realizations $X'_{\ell_i} \in \mathcal{X}_{\ell_i}$.

We represent the legal ecosystem at time t for jurisdiction i as a structured knowledge state $\mathcal{L}_t^{(i)}$. This knowledge state includes both EU-level and national-level instruments and their dependencies, and it evolves over time as amendments, new acts, repeals, and new jurisprudential interpretations occur.

 doi:10.20944/preprints202603.1216.v1

13.3. Legal Knowledge State as a Normative Graph

For each jurisdiction i and time t, define the normative graph

$$\mathcal{G}_t^{(i)} = (V_t^{(i)}, E_t^{(i)}), \tag{64}$$

where nodes $V_t^{(i)}$ correspond to legal instruments, concepts, or clauses, and edges $E_t^{(i)}$ encode typed dependencies such as:

- *hierarchy* (constitutional/primary/secondary law),
- *incorporation* (EU act transposition, direct applicability),
- *reference* (cross-references, definitional pointers),
- *override* (lex specialis, lex posterior, explicit derogations),
- *scope linkage* (domain applicability constraints).

The legal knowledge state $\mathcal{L}_t^{(i)}$ comprises $\mathcal{G}_t^{(i)}$ together with version metadata, canonical terminology inventories, and interpretive constraints (e.g., stable definitions and drafting conventions).

13.4. Structural Normalization for Normative Texts

We introduce a domain-specific normalization operator for legal discourse,

$$\mathcal{N}_{\text{legal}} : \mathcal{X}_\ell \to \mathcal{Z}_{\text{legal}}, \tag{65}$$

mapping a surface legal text into a structural representation in a legal structural space $\mathcal{Z}_{\text{legal}}$.

For a source proposal X, define

$$Z_{\text{legal}} = \mathcal{N}_{\text{legal}}(X). \tag{66}$$

A compact structural state sufficient for normative preservation is

$$Z_{\text{legal}} = (D, O, E, S, R), \tag{67}$$

where:

- D encodes definitional content (terms, definitions, concept identifiers),
- O encodes obligations/prohibitions/permissions (deontic operators + subjects + actions),
- E encodes exceptions and derogations (conditions under which O is modified),
- S encodes sanctions and enforcement mechanisms (where applicable),
- R encodes references and dependency hooks (explicit cross-references and implicit alignments to external instruments).

This representation emphasizes normative meaning as a structured dependency object rather than as a lexical string.

13.5. Pivot-Language Anchoring for Canonical Terminology

Many domains contain entities for which surface words in different languages are incomplete or ambiguous identifiers. Operationally, translation quality improves when lexical realization is mediated by a canonical "pivot" identifier that is stable across languages, and then projected into the target language through a controlled mapping.

We formalize this mechanism by introducing, for each domain, a canonical identifier function

$$\gamma_{\mathcal{D}} : \text{Entity} \to \Gamma_{\mathcal{D}}, \tag{68}$$

where $\Gamma_{\mathcal{D}}$ is a space of canonical identifiers.

For a tokenized mention τ in a source text, the pipeline extracts a pivot identifier

$$\pi(\tau) = \gamma_{\mathcal{D}}(\tau), \tag{69}$$

 doi:10.20944/preprints202603.1216.v1

and constrains realization in the target language to preserve $\pi(\tau)$ under an allowed lexicalization map.

This "pivot bridge" generalizes beyond legal discourse and admits high-value specializations:

Biology (Latin binomials).

For taxa (animals, plants, microorganisms), canonical identity is captured by Latin binomial nomenclature. A Bulgarian mention (e.g., "врабче") is first mapped to a Latin pivot (e.g., *Passer domesticus* when context indicates the house sparrow), and then realized in English as *sparrow* with optional taxonomic specificity. This avoids underspecified realizations (e.g., "little bird") and supports disambiguation when multiple common names exist.

Chemistry (systematic nomenclature).

Chemical entities admit stable canonical identity via systematic names and formula-level identifiers. A source-language mention is normalized to a canonical chemical pivot (e.g., IUPAC-style systematic name, molecular formula, or structured identifier), and target realization is constrained to reflect the same chemical identity. This supports translation under synonyms, trivial names, and local conventions.

Medicine and biology (Latin/Greek term inventories).

Clinical and anatomical terminology often contains Latin or Greek-based canonical forms that function as stable anchors. A pivot identifier supports consistent cross-language realization of organs, pathologies, and procedures with reduced terminological drift.

Music (Italian performance terms).

Musical instructions frequently use Italian canonical terms (e.g., *allegro*, *adagio*, *crescendo*). Translation preserves the canonical Italian pivot while permitting target-language paraphrase only when style guidelines require it. This yields faithful realization of tempo and articulation semantics while retaining domain conventions.

Law (Latin legal terms).

Legal discourse contains Latin terms with stable doctrinal meaning (e.g., *lex specialis*, *mens rea*, *habeas corpus*). A pivot representation preserves the canonical Latin anchor and constrains jurisdictional lexicalization to appropriate target formulations.

In all such domains, pivot anchoring becomes a structural component that can be validated and version-controlled. The canonical inventory itself is time-indexed: new scientific consensus, terminology standards, or interpretive guidance may update the pivot mapping without changing the source text.

13.6. Jurisdiction-Conditioned Constraint Space

For a target jurisdiction i at time t, define a constraint set

$$\mathcal{C}_t^{(i)} = \mathcal{C}_{\mathrm{EU},t} \cup \mathcal{C}_{\mathrm{Nat},t}^{(i)} \cup \mathcal{C}_{\mathrm{Draft}}^{(i)}, \tag{70}$$

where:

- $\mathcal{C}_{\mathrm{EU},t}$ encodes compatibility with the current EU acquis (definitions, scope, mandatory formulations, reference forms),
- $\mathcal{C}_{\mathrm{Nat},t}^{(i)}$ encodes compatibility with the current national law of jurisdiction i (terminology, institutional names, procedural constraints, legal concepts),
- $\mathcal{C}_{\mathrm{Draft}}^{(i)}$ encodes drafting conventions and style guides for jurisdiction i (shall/must usage, article structure, numbering, phraseology).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

Crucially, $\mathcal{C}_t^{(i)}$ is time-indexed: it changes even when X is unchanged. When pivot-anchored terminology is present, $\mathcal{C}_t^{(i)}$ also includes canonical terminology constraints ensuring that each pivot $\pi(\tau)$ is realized through an admissible target lexicalization and institutional naming convention.

13.7. Reconstruction Objective

For each jurisdiction i and time t, the target-language realization is defined as a constrained reconstruction:

$$X_t'^{(i)} = \arg\min_{Y \in \mathcal{X}_{\ell_i}} \left[D_{\text{legal}}\left(\mathcal{N}_{\text{legal}}(Y), Z_{\text{legal}} \mid \mathcal{L}_t^{(i)}\right) + \lambda\, \mathcal{L}_{\mathcal{C}_t^{(i)}}(Y) \right], \tag{71}$$

where $D_{\text{legal}}(\cdot,\cdot \mid \mathcal{L}_t^{(i)})$ measures structural deviation in the presence of the legal knowledge state, and $\mathcal{L}_{\mathcal{C}_t^{(i)}}(Y)$ is a constraint-violation functional (hard constraints can be implemented as infinite penalty or feasibility constraints).

A legally oriented decomposition of the structural divergence is:

$$D_{\text{legal}} = \alpha_1 D_D + \alpha_2 D_O + \alpha_3 D_E + \alpha_4 D_S + \alpha_5 D_R, \tag{72}$$

where each term compares the corresponding component of Z_{legal} between source and candidate, including preservation of reference structure and definitional anchoring. When pivot identifiers are used, D_D and D_R also include pivot-consistency deviation, penalizing realizations that drift away from canonical identifiers.

13.8. Versioning and Time Dependence

We treat the translation as a version-controlled artifact. Let $\text{ver}(X)$ denote the source document version, and let $\text{ver}(\mathcal{L}_t^{(i)})$ denote the version state of the legal ecosystem relevant to jurisdiction i. A translation instance is indexed as

$$\text{Trans}(X, i, t) = X_t'^{(i)}, \tag{73}$$

and it is valid relative to the pair $(\text{ver}(X), \text{ver}(\mathcal{L}_t^{(i)}))$. This captures the central operational point: even with fixed $\text{ver}(X)$, updates in $\text{ver}(\mathcal{L}_t^{(i)})$ may require a new $X_{t+1}'^{(i)}$.

The same indexing principle applies to canonical pivot inventories: updates in a terminology standard or authoritative mapping (e.g., revisions in a taxonomy database, an updated chemical nomenclature guideline, or an updated legal terminology inventory) may trigger revalidation even when both $\text{ver}(X)$ and core legal instruments are unchanged.

13.9. Legal Update as Ecosystem Update

When the legal ecosystem changes,

$$\mathcal{L}_t^{(i)} \to \mathcal{L}_{t+1}^{(i)}, \tag{74}$$

the constraint set updates as

$$\mathcal{C}_t^{(i)} \to \mathcal{C}_{t+1}^{(i)}. \tag{75}$$

This update is operationally distinct from the poetic epistemic update. Here the driver is not interpretive evidence about an author, but authoritative changes in binding instruments, interpretive acts, reference structures, and canonical terminology inventories. The system therefore implements deterministic, auditable update logic and strict version control.

13.10. Change Propagation and Impact Analysis

Given a legal update, the system identifies an affected subgraph:

$$\Delta\mathcal{G}_{t\to t+1}^{(i)} \subseteq \mathcal{G}_{t+1}^{(i)}, \tag{76}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1216.v1

and computes the set of translations impacted by this change:

$$\mathcal{T}_{\text{imp}}^{(i)} = \left\{ X_{t'}^{\prime(i)} \ : \ \text{Dep}\left(X_{t'}^{\prime(i)}\right) \cap \Delta\mathcal{G}_{t\to t+1}^{(i)} \neq \emptyset \right\}, \tag{77}$$

where $\text{Dep}(\cdot)$ returns the set of dependency hooks (definitions, references, invoked procedures) used by a translation.

If pivot-anchored terminology is used, $\text{Dep}(\cdot)$ also includes the set of canonical pivots invoked by the translation. This enables impact analysis under updates in canonical inventories (e.g., when a taxonomic pivot is refined, a chemical identifier mapping changes, or a legal terminology inventory is revised).

Each impacted translation is scheduled for revalidation and, if necessary, reconstruction under the updated constraints.

13.11. Agentic Architecture over a Generative Model

The framework admits an agentic pipeline that uses a large language model as a realization engine and enforces legal correctness via external structural and constraint layers.

We define the modules:

- **Legal Structural Extractor** $\mathcal{E}_{\text{legal}}$: computes $Z_{\text{legal}} = \mathcal{N}_{\text{legal}}(X)$ and extracts dependency hooks, including canonical pivots where applicable.
- **Legal Knowledge Manager** $\mathcal{K}$: maintains $\mathcal{L}_t^{(i)}$ and computes updates $\mathcal{L}_t^{(i)} \to \mathcal{L}_{t+1}^{(i)}$ with version provenance.
- **Constraint Builder** $\mathcal{M}_{\text{legal}}$: constructs $\mathcal{C}_t^{(i)}$ from $\mathcal{L}_t^{(i)}$, drafting conventions, and canonical terminology inventories.
- **Generative Realizer** $\mathcal{G}$: produces candidate realizations Y in ℓ_i conditioned on $(Z_{\text{legal}}, \mathcal{C}_t^{(i)})$, including pivot-consistent lexicalization.
- **Validator** $\mathcal{V}_{\text{legal}}$: checks definitional alignment, deontic preservation, reference integrity, pivot consistency, and jurisdictional admissibility; returns diagnostics for repair.
- **Version Controller** $\mathcal{U}$: assigns translation identifiers, stores provenance, and enforces auditability and rollback.

The operational loop is generate $\to$ validate $\to$ repair, with strict logging of constraint violations and corrections.

13.12. Evaluation Protocol

A legally meaningful evaluation emphasizes:

- **Normative equivalence:** preservation of obligations, exceptions, and sanctions under the target legal semantics.
- **Reference integrity:** correctness of cross-references and definitional anchoring.
- **Pivot consistency:** preservation of canonical identifiers where domain terminology admits pivot anchoring.
- **Jurisdictional admissibility:** compatibility with national drafting conventions and institutional naming.
- **Update robustness:** correct propagation under changes in $\mathcal{L}_t^{(i)}$ or canonical inventories without source edits.
- **Auditability:** ability to trace each translation decision to versions of $(X, \mathcal{L}_t^{(i)})$ and constraint sets.

13.13. Interpretation

Normative translation is formulated as jurisdiction-conditioned structural reconstruction within an evolving legal ecosystem. The central technical challenge is not lexical substitution but preservation of normative structure under time-indexed constraints induced by a changing web of EU and national instruments.

 doi:10.20944/preprints202603.1216.v1

The pivot-anchoring mechanism generalizes this perspective to other terminology-heavy domains (biology, chemistry, medicine, and music), where canonical identifiers act as stable cross-linguistic anchors. In all such settings, translation becomes a structured reconstruction problem: structural invariants are preserved while surface realization is constrained by a time-indexed ecosystem of authoritative instruments, standards, and canonical inventories.

This framing naturally yields version-controlled translation artifacts whose updates are triggered both by source revisions and by ecosystem-level changes in referenced, governing, or canonically defining resources.

References

1. Gentner, D. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* **1983**, *7*, 155–170.
2. Falkenhainer, B.; Forbus, K.D.; Gentner, D. The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence* **1989**, *41*, 1–63.
3. Holyoak, K.J.; Thagard, P. *Mental Leaps: Analogy in Creative Thought*; MIT Press, 1995.
4. Hokamp, C.; Liu, Q. Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search. In Proceedings of the Proceedings of ACL, 2017, pp. 1535–1546.
5. Anderson, P.; Fernando, B.; Johnson, M.; Gould, S. Guided Open Vocabulary Image Captioning with Constrained Beam Search. In Proceedings of the Proceedings of EMNLP, 2017, pp. 936–945.
6. Lu, X.; Welleck, S.; Hessel, J.; Choi, Y. NeuroLogic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints. In Proceedings of the Proceedings of NAACL, 2022, pp. 2373–2387.
7. Dathathri, S.; Madotto, A.; Lan, J.; Hung, J.; Frank, E.; Molino, P.; Yosinski, J.; Liu, R. Plug and Play Language Models: A Simple Approach to Controlled Text Generation. In Proceedings of the Proceedings of ICLR, 2020.
8. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* **2020**, *33*, 1877–1901.
9. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training Language Models to Follow Instructions with Human Feedback. *Advances in Neural Information Processing Systems* **2022**, *35*, 27730–27744.
10. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971* **2023**.
11. OpenAI. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* **2023**.
12. Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, *27*, 379–423, 623–656.
13. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Annals of Mathematical Statistics* **1951**, *22*, 79–86.
14. Lin, J. Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory* **1991**, *37*, 145–151.
15. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.
16. Conneau, A.; Lample, G.; Ranzato, M.; Denoyer, L.; Jégou, H. Word Translation Without Parallel Data. In Proceedings of the Proceedings of ICLR, 2018.
17. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Proceedings of NAACL, 2019, pp. 4171–4186.
18. Ruder, S.; Vulić, I.; Søgaard, A. A Survey of Cross-Lingual Word Embedding Models. *Journal of Artificial Intelligence Research* **2019**, *65*, 569–631.
19. Alexandrov, A.; Raychev, V.; Dimitrov, D.I.; Zhang, C.; Vechev, M.; Toutanova, K. BgGPT 1.0: Extending English-centric LLMs to other languages. *arXiv preprint arXiv:2412.10893* **2024**, [arXiv:cs.CL/2412.10893].