Computational Analysis of Contemporary Romanian Poetry

Anonymous ACL submission

Abstract

This paper presents a computational analysis of contemporary Romanian poetry, targeting fine-grained enjambment detection and macro-thematic classification. We introduce **RoPoetry2K**, a novel, human-annotated corpus capturing the stylistic and semantic nuances of the post-2000 literary landscape. Our methodology enhances a native RoBERT encoder via Domain-Adaptive Pre-Training, sliding-window chunking for long-form verse, and a hybrid weak-supervision regime combining programmatic silver labels with gold data. While frontier LLMs like Gemini Pro 3.1 demonstrate strong generalist capabilities in mapping abstract global themes, our specialized pipelines systematically outperform them in decoding complex syntactic ruptures and localized cultural niches (e.g., *Love & Intimacy & Mundane Life*). Ultimately, we demonstrate that resolving syntax-semantic mapping challenges in high-entropy verse necessitates a mixed approach; by ensembling specialized BERT-like encoders with general-purpose LLMs, we achieve state-of-the-art performance, providing a robust computational poetics framework seamlessly exportable to both low-resource and widely spoken languages.

1 Introduction

"Can a machine hear the *silence* within a poem?" When asked about technology's impact on literature, acclaimed Romanian poet Nora Iuga remarked: "The only definition of great poetry is silence" (Iuga, 2019). This ontological paradox challenges Natural Language Processing (NLP). If poetic essence lies in the unsaid—the deliberate silence between lines—how can computational models, inherently bound to the statistical surface of tokens, decode it?

Computationally, modern free verse is a high-entropy environment defined by non-linear syntax, dense ellipsis, and structural fragmentation. One primary driver of this fragmentation is *enjambment*—a deliberate syntactic rupture across line breaks. Much like syncopation in jazz music, enjambment disrupts the expected structural rhythm, inserting a micro-pause or "short silence" that generates cognitive completion tension. Standard language models, pre-trained primarily on normative prose, systematically struggle to process these aesthetic domain shifts (Dipper and Laarmann-Quante, 2024; Miani et al., 2025). To computationally detect these structural ruptures, NLP architectures must move beyond traditional sentence parsing to explicit sequence-pair boundary modeling, evaluating the grammatical tension strictly at the junction between consecutive lines. Using contemporary Romanian verse as a morphosyntactically flexible testbed, we demonstrate how localized, domain-specific training dynamics can overcome the limitations of general-purpose frontier models.

To address the challenge of "hearing the silence," we present a specialized NLP pipeline targeting macro-thematic classification and fine-grained enjambment detection. Our contributions are three-fold: (**i**) We introduce **RoPoetry2K**, a human-annotated corpus capturing the stylistic and semantic nuances of post-2000 Romanian poetry; (**ii**) We propose a RoBERT-based architecture enhanced with Domain-Adaptive Pre-Training (DAPT) (Gururangan et al., 2020) for thematic mapping, alongside a hybrid programmatic weak-supervision regime for enjambment detection; (**iii**) We establish a rigorous benchmark against Gemini Pro 3.1 and demonstrate that computational poetry analysis on such datasets necessitates a mixed approach. By combining the global semantic reach of generalist LLMs with the precise syntactic mapping of specialized BERT-like encoders via a late-fusion ensemble, we achieve state-of-the-art performance.

2 Related Work

Early Romanian structuralist foundations in mathematical linguistics formalized poetic verse through information theory, modeling rhythmic features as non-linear misalignments between syntactic flows and metrical boundaries (Marcus, 1970; Dinu, 1986). In highly inflected languages like Romanian, this structural instability is further amplified by irregular lexical patterns, demonstrating that deterministic, rule-based templates are insufficient and highlighting the necessity for flexible, data-driven computational approaches (Ciobanu et al., 2014).

The structural divergence between poetry and prose poses cross-lingual parsing challenges requiring specialized linguistic processing. For instance, applying Universal Dependencies (UD) to German verse highlights severe difficulties in sentence boundary recognition and the parsing of atypical syntactic constructions (Dipper and Laarmann-Quante, 2024). Similarly, prose-trained models fail to generalize to Old English poetic syntax without targeted adaptation (Miani et al., 2025). To mitigate the absence of dedicated poetic treebanks, derived noisy syntax trees have successfully enhanced downstream thematic classification in ancient Chinese (Wang et al., 2023). Furthermore, capturing the aesthetic essence of poetry often requires custom feature engineering beyond standard word embeddings, as seen in the computational classification of phonetic and syntactic attributes in Sanskrit verse (Barbadikar and Kulkarni, 2024). Building on this need for specialized architectures, recent advancements demonstrate that domain-specific representation learning is critical; for instance, Huang and Shen (2025) introduced PoemBERT, proving that dynamic masking strategies and custom phonetic embeddings outperform standard language models in capturing the unique imagery of classical verse. Beyond localized feature extraction, tracking the stylistic evolution of poetry across temporal boundaries introduces distinct classification and evaluation complexities; Fatima et al. (2026) developed a diachronic Urdu poetry corpus to classify segments across classical, modern, and contemporary eras, emphasizing that strict author-disjoint evaluation strategies are critical to prevent neural models from overfitting to individual poetic idiosyncrasies rather than capturing genuine generational styles.

Consequently, computational frameworks have shifted toward hybrid setups to capture the multi-constrained nature of poetic text, though purely statistical generative methods historically struggle to encapsulate deep semantic layers alongside formal aesthetic constraints (Hopkins and Kiela, 2017). This limitation persists in frontier Large Language Models (LLMs); despite high surface fluency, they exhibit systemic stylistic deficiencies, producing highly repetitive outputs regardless of prompt variations (Chen et al., 2024b). Beyond lexical diversity, LLMs frequently fail to grasp the spatial and visual semantics of poetry; recent research demonstrates that LLMs struggle to appropriately generate or interpret whitespace—a critical structural feature that dictates poetic rhythm and form—highlighting the gap between sequential token processing and visual layout comprehension (Bhyravajjula et al., 2025).

2.1 Computational Enjambment Detection

Rule-Based Approaches. Ruiz Fabo et al. (2017) ground their rule-based framework in the theories of Quilis (1964), formalizing enjambment as a mismatch between syntactic and metrical units. Utilizing the IXA pipeline (Agerri et al., 2014) for dependency parsing, their system applies approximately 30 heuristic rules to identify severed syntactic units across line breaks in Spanish sonnets. While highly accurate on contemporary texts, parsing errors severely degrade performance on diachronic corpora due to archaic language and *hyperbaton* (syntactic displacement).

Deep Learning Approaches. Conversely, Baumann et al. (2018) tackled the complexities of German free verse using a multimodal hierarchical attention network (HAN) that integrates character-level text encoding with acoustic and prosodic features. Although the network achieved high enjambment detection accuracy (F_1: 0.91), it did not naturally prioritize these structural breaks when classifying downstream poetic styles. Instead, attention weights were distributed evenly across all verses, suggesting that for neural architectures, stylistic identity is processed holistically rather than being strictly localized at enjambment sites.

3 Computational Formalization of Enjambment

The visual layout of poetry fundamentally shapes its cognitive processing. Koops van' t Jagt et al. (2014) define enjambment as a structural conflict where a line break severs a syntactic boundary. Eye-

Oamenii au decis să înregistreze un videoclip (1)
['The people decided to record a video']
în care vorbesc despre ei și despre planeta lor. (2)
['in which they talk about themselves and about their planet.']
Dar nu știau că cei care vizionează videoclipul (3)
['But they did not know that those who watch the video']
nu sunt prietenoși. (4)
['are not friendly.']

Figure 1: Illustration of enjambment extracted from Poem 1015. The visual line breaks sever tight syntactic dependencies, such as clausal junction (`cross_clause`) between verses 1-2 and subject-verb relation (`subj_verb`) between 3-4.

tracking empirical studies demonstrate that these ruptures trigger distinct reading patterns—such as slower rates and higher regression frequencies—compared to intact verses (see Figure 1).

Computationally, we formalize enjambment detection as a granular sequence-pair boundary classification task. Rather than treating line breaks as standard sentence delimiters, our framework models the textual junction between consecutive verses to identify severed token-level dependencies.

3.1 Syntactic Cohesion and Hierarchical Typology

Successful computational approaches rely on the concept of *syntactic cohesion* (Ruiz Fabo et al., 2017). Through auditory perception experiments, Quilis (1964) established that certain cohesive word groups resist pauses in natural speech; versification constraints rupturing these groups generate the stylistic effect of enjambment.

Adapting the classical frameworks of Quilis (1964) and Spang (1983), we define a hierarchical typology tailored to the flexible word order and morphological complexity of Romanian. We categorize these cross-line syntactic ruptures into four macro-groups (comprehensive definitions and validated corpus examples are detailed in Appendix A.4):

- **Lexical enjambment:** A word divided across lines via a trailing hyphen.
- **Phrase-bounded enjambment:** Breaks within a single phrase, disrupting closely-knit links such as adjective-noun (`pb_adj_noun`), verb-adverb (`pb_verb_adv`), or core prepositions.
- **Cross-clause enjambment:** Line breaks occurring immediately before coordinating or subordinating conjunctions.
- **Expansion enjambment:** Ruptures separating clausal elements like subject-verb or verb-object, traditionally considered looser bonds than intra-phrase links.

Because this taxonomy relies on cross-lingual dependency relations rather than language-specific morphology, it is highly exportable. Although tested on a flexible-syntax language, this heuristic structure can be directly mapped onto widely spoken languages (e.g., English) requiring only minimal alignment to standard POS and dependency tags.

4 RoPoetry2K: The Romanian Contemporary Poetry Corpus

4.1 Corpus Description

RoPoetry2K contains 1,028 contemporary Romanian poems written by 69 unique authors from the post-2000 period. To provide a holistic computational evaluation of both poetic form and content, we annotated the corpus along two distinct axes, defining our core research tasks: **Task 1** targets micro-syntactic enjambment detection, while **Task 2** focuses on macro-thematic classification. In contemporary free verse, where traditional metric constraints are frequently abandoned, enjambment emerges as the primary structural device governing internal rhythm (Tașcu, 2006). Paired with thematic labeling—which maps the semantic landscape of the post-2000 generation—these tasks establish a robust benchmark testing an NLP model's capacity to navigate both granular syntactic disruptions and high-level abstractions.

We established the ground truth for both tasks through a structured two-phase annotation pipeline. First, a multi-expert pilot phase was conducted to calibrate the annotation guidelines and validate Inter-Rater Agreement (IRA). Subsequently, a lead domain expert completed the task-specific annotations, curating the targeted 107-poem Gold Standard for enjambment (Task 1) and labeling the full RoPoetry2K corpus for macro-themes (Task 2). Complete annotator demographics, IRA metrics, and adjudication protocols are detailed in Appendix A.1.

The texts were harvested from prominent digital archives and anthologies (detailed in Appendix A), yielding 177,480 tokens and a high Type-Token Ratio (TTR) of 0.1523. Poem lengths average 172.65

words ($\pm$215.73 SD), reflecting the structural fluidity of modern poetry, with texts ranging from 8 to 5,137 words. We observe a similar variation in the number of lines per poem, with an average of 27.22 lines ($\pm$33.22 SD) and a median of 22 lines. To ensure the dataset represents generational stylistic variance rather than individual idiosyncrasies, the collection is strictly author-balanced; the top five contributors account for only 31.5% of the total dataset, with the remainder evenly distributed among 64 other poets.

4.2 Task 1: Enjambment Annotation and Gold Standard

Manual annotation of enjambment is a meticulous process requiring careful linguistic analysis at every line break. To map this phenomenon, we initially performed complete, line-by-line scansion of 50 fully intact poems to capture the natural distribution of enjambments within text units.

However, an evaluation of this initial set revealed severe class imbalance. While common structural breaks appeared frequently, niche syntactic categories (e.g., `pb_adv_prep` and `pb_adj_adv`) contained fewer than 20 examples, making them too rare for reliable model evaluation. To resolve this imbalance, we conducted targeted sampling across the remaining corpus, extracting and annotating isolated line pairs containing these rare syntactic constructions. This targeted enrichment expanded our final Gold Standard to a total of 107 unique poems, comprising 1,537 annotated line pairs.

The dataset contains an average of 14.36 line pairs per poem. As detailed in the comprehensive taxonomy in Appendix A.4, the majority of line transitions do not feature a syntactic break, with the *NONE* class accounting for 70.85% of the data (1,089 instances). Among the actual enjambment types, cross-clause breaks (5.47%) and relative-word disruptions (5.34%) represent the most frequent structural categories, providing a robust baseline for evaluation.

4.3 Task 2: Thematic Labeling

The initial thematic annotation phase utilized a detailed multi-label framework comprising eight distinct themes to capture the modern lyrical landscape (Lupașcu, 2022). While this approach captured thematic density (averaging 2.95 labels per poem), it introduced substantial semantic noise and severe class imbalance unsuitable for standard text classification benchmarks.

To establish a robust, mutually exclusive single-label framework, we executed a data consolidation strategy mapping low-frequency and overlapping categories into five consolidated macro-classes (e.g., merging *Love & Intimacy* with *Mundane & Daily Life*, and integrating *Eco-Consciousness* into *Digital Realities & Posthumanism*). Primary dominant themes were resolved via expert consensus, yielding a balanced dataset optimized for machine learning evaluation (detailed validation semantics are provided in Appendix A.2). Table 1 outlines the final class distribution.

Consolidated Theme	Count	%
Love & Intimacy & Mundane Life	334	32.49%
Digital Realities & Posthumanism	208	20.23%
Identity & Otherness	188	18.28%
Self-Referentiality (Ars Poetica)	152	14.78%
Philosophical & Ontological Inquiry	146	14.20%

Table 1: Final consolidated thematic class distribution in the corpus.

5 Methodology

5.1 Programmatic Distant Supervision

While manual annotation provides a high-quality evaluation set, its size limits the capacity of deep neural networks to learn generalized patterns for complex multi-class topologies. To scale up our training data without the prohibitive cost of manual labeling, we implemented a distant supervision pipeline to generate a massive silver dataset from the remaining unannotated corpus. This pipeline operationalizes the syntactic rules derived from our methodology (Section 3.1) into a programmatic labeling function (lf_spacy). Built upon the `ro_core_news_lg` pipeline from spaCy (Honnibal et al., 2020), the function inspects the boundary between consecutive line pairs.

It analyzes part-of-speech (POS) tags, dependency relations (e.g., `amod`, `nsubj`), and morphological attributes to programmatically detect structural breaks. A critical constraint of this automated approach is managing noise. If a line break does not explicitly trigger a structural rule, or if the syntactic context is ambiguous, the function returns an `ABSTAIN` flag. All `ABSTAIN` instances are systematically filtered out, ensuring the neural models learn only from high-confidence supervision signals. Empirical evaluation of lf_spacy against our manual annotations confirms its viability: it yields a rule coverage of 42.55% while maintaining a high

weighted average precision of 85.0% on the Gold Standard (a detailed confusion matrix and FP/FN analysis are provided in Appendix A.3). This programmatic approach successfully expanded our data to 10,072 clean, syntactically-grounded training examples.

5.2 Specialized Language Modeling Architecture

For our specialized modeling pipeline, we utilize **RoBERT-base** (Masala et al., 2020), optimizing its architecture specifically for our dual tasks.

For Task 1, consecutive verses are joined into a single sequence (`"Line N [SEP] Line N+1"`) and mapped to either a flat 14-class topology or a multi-stage hierarchical cascade. To handle the severe class imbalances present in the poetic datasets, we extended the network's default objective with a custom weighted cross-entropy loss, dynamically adjusting gradients via inverse class frequencies. For Task 2, the base encoder first undergoes Domain-Adaptive Pre-training (DAPT) (Gururangan et al., 2020) via Masked Language Modeling on our unannotated poetry corpus, adapting its internal representations to modern literary morphology before appending a linear classification head. Detailed hyperparameters, architectural dimensions, and optimization callbacks for all experiments are provided in Appendix A.6.

5.3 Evaluation Protocols

To evaluate the efficacy of our proposed methodology, we contrast our domain-adapted pipeline against a general-purpose frontier LLM across both tasks.

For Task 1, we employ 5-fold cross-validation using a `GroupKFold` split strategy based on `poem_id`. This strict constraint ensures that all line pairs originating from the same poem are kept entirely within a single fold, completely preventing cross-fold data leakage.

For Task 2, the corpus is partitioned into an 80% training split and a 20% stratified test split. Because full-text contemporary poems frequently exceed the native sequence boundaries of transformer encoders, the model processes documents via a sliding window text-chunking strategy. During inference, text-level predictions are resolved by flattening the probability distributions of all overlapping chunks via a max-pooling operation, followed by the application of class-weight multipliers optimized via grid-search to balance minority themes.

5.4 LLM Baseline Using Gemini Pro 3.1

To establish a competitive general-purpose baseline, we evaluated Gemini Pro 3.1 [1] (Gemini Team et al., 2023) in a strict zero-shot regime. To mitigate the structural ambiguity of contemporary verse and ensure strict parsing reliability, we designed structured prompts tailored to each task. For Task 1, the prompt utilizes an expert philologist persona, formal linguistic definitions, Chain-of-Thought (CoT) reasoning to systematically isolate sequence boundaries, and strict JSON output constraints. For Task 2, the model assumes a literary critic persona, evaluating the full text against the explicit boundaries of the 5 consolidated macro-themes. The full execution prompts are provided in Appendix A.7.

6 Results and Findings

6.1 Task 1: Enjambment Detection Results

To evaluate the impact of structural design, programmatic weak supervision, and cross-architecture collaboration, we deploy our enjambment pipeline across five configurations: **(1) Pure Gold Standard:** The baseline RoBERT model trained exclusively on manually annotated gold line pairs; **(2) Pure Cascading:** A two-stage hierarchical model where a binary gate filters `NONE` instances before passing candidates to a 13-class specialist; **(3) Hybrid Flat:** A single-stage 14-class model pre-trained on the programmatic silver dataset and fine-tuned on gold; **(4) Cascading Hybrid:** A framework combining the hierarchical cascade with dual-stage silver pre-training; and **(5) Ensemble:** A heuristic routing model that leverages the complementary strengths of both our specialized Hybrid Flat encoder and Gemini Pro 3.1.

Table 2 summarizes the fine-grained performance (F_1-score) across all 14 syntactic categories, positioning our five specialized configurations against the zero-shot Gemini Pro 3.1 baseline.

The empirical results reveal that training exclusively on the limited gold dataset (*Pure Gold*) establishes a constrained baseline (macro F_1: 0.511) due to severe data starvation among minority syntactic classes. This bottleneck is highly visible in the hierarchical *Pure Cascading* model, where the macro F_1-score drops to its lowest point (0.458). Isolating the 13 enjambment types into a secondary multiclass tier severely reduces the active training

[1] Model card available at: `https://deepmind.google/models/model-cards/gemini-3-1-pro/`.

Category	Pure Gold	Cascading	Hybrid Flat	Cascading Hybrid	Gemini Pro 3.1	Ensemble
NONE	0.879	0.899	0.863	0.830	0.862	0.900
lexical	0.775	0.787	0.800	0.800	1.000	0.970
pb_adj_adv	0.400	0.493	0.602	0.571	0.687	0.690
pb_adj_noun	0.358	0.257	0.596	0.531	0.743	0.630
pb_adj_prep	0.580	0.437	0.548	0.479	0.641	0.700
pb_adv_prep	0.441	0.510	0.831	0.691	0.800	0.730
pb_noun_prep	0.270	0.191	0.561	0.494	0.558	0.580
pb_verb_adv	0.311	0.339	0.544	0.521	0.500	0.530
pb_verb_chain	0.844	0.494	0.865	0.767	0.927	0.930
pb_verb_cprep	0.357	0.160	0.560	0.527	0.536	0.690
pb_relword	0.685	0.737	0.842	0.784	0.633	0.900
cross_clause	0.357	0.306	0.439	0.440	0.270	0.460
subj_verb	0.253	0.252	0.580	0.366	0.750	0.710
obj_verb	0.652	0.551	0.787	0.602	0.849	0.880
Macro F_1	**0.511**	**0.458**	**0.673**	**0.600**	**0.697**	**0.740**

Table 2: Performance comparison (F_1-scores) for fine-grained enjambment detection on the Gold Standard.

examples per class, causing catastrophic drops in categories like pb_verb_cprep (F_1: 0.160).

Injecting programmatic distant supervision completely reverses this trend. Our proposed *Cascading Hybrid* setup mitigates the data scarcity of the pure hierarchy, boosting the macro F_1-score by over 14% to 0.600. This confirms that hierarchical architectures require large-scale weak supervision to align multi-stage decision boundaries. However, due to slight error propagation at the initial binary gate (where false negatives cannot be recovered downstream), the single-stage *Hybrid Flat* model emerges as our absolute specialized champion, achieving a macro F_1 of 0.673.

While the massive scale of Gemini Pro 3.1 secures the highest overall baseline (macro F_1: 0.697), our specialized, domain-aware models systematically outperform the frontier LLM on complex native syntactic disruptions.

On cross_clause boundaries, Gemini's performance drops to a poor F_1 of 0.270. In contrast, both our *Hybrid Flat* and *Cascading Hybrid* models maintain a stable F_1 of 0.439, capturing the cross-line dependency that the LLM misinterprets as a standard prose sentence transition. Similarly, for relative-word disruptions (pb_relword), our *Hybrid Flat* model achieves an outstanding F_1 of 0.842 compared to Gemini's 0.633, driven by our native spaCy programmatic rules.

To leverage these complementary architectural strengths, we implemented a rule-based *Ensemble*. By heuristically routing predictions—trusting RoBERT for fine-grained functional dependencies (where it acts as a domain expert) and Gemini for broader semantic disruptions—the ensemble correctly resolves the individual blind spots of both models. Crucially, this routing strategy is not an artifact of test-set overfitting, but an *a priori* linguistically motivated rule: all categories representing purely syntactic or functional severing (e.g., prepositions, relative pronouns) are statically routed to the specialized encoder. This routing strategy (detailed in Appendix A.8) establishes a new state-of-the-art for our dataset, pushing the overall Macro F_1 to an impressive 0.740 and proving that smaller, targeted encoders can be seamlessly integrated with global generative giants to process high-entropy literary syntax.

6.2 Task 2: Thematic Classification Results

In our second task, thematic classification, we evaluate whether our domain-adapted RoBERT can match or surpass the semantic mapping capabilities of a generalist LLM across contemporary poetic macro-themes.

Table 3 centralizes the performance metrics across all 5 consolidated macro-themes, establishing a direct comparison between Gemini Pro 3.1 and our fine-tuned RoBERT architecture.

Theme	Gemini (F_1)	RoBERT (F_1)
Digital Realities...	0.633	0.550
Identity...	0.500	0.580
Love & Mundane...	0.537	0.760
Philosophical...	0.531	0.330
Self-Referentiality...	0.425	0.540
Macro Precision	0.541	0.580
Macro Recall	0.532	0.550
Macro F_1	**0.525**	**0.550**

Table 3: Thematic classification performance (Gemini Pro 3.1 vs. RoBERT).

The quantitative evaluation demonstrates the clear benefit of specialized domain adaptation, with our optimized RoBERT model achieving a higher overall baseline (Macro F_1: 0.550) compared to Gemini Pro 3.1 (Macro F_1: 0.525).

A granular class analysis highlights a sharp divergence in architectural strengths. Our fine-tuned RoBERT model completely dominates the majority macro-theme, *Love & Intimacy & Mundane Life*, reaching an outstanding F_1-score of 0.760 against Gemini's 0.537. This is a direct consequence of our post-processing prediction multipliers and window-aggregation strategy, which successfully align the encoder with localized cultural and domestic markers typical of post-2000 Romanian realism. RoBERT also retains a clear edge in capturing fluid identity dynamics (*Identity & Otherness*, F_1: 0.580 vs 0.500) and meta-poetic structures (*Self-Referentiality*, F_1: 0.540 vs 0.425).

Conversely, Gemini Pro 3.1 exhibits superior performance in abstract or conceptually uniform axes. It achieves a significantly higher score in *Philosophical & Ontological Inquiry* (F_1: 0.531 vs 0.330) and leads in *Digital Realities & Posthumanism* (F_1: 0.633 vs 0.550). This behavioral split occurs because massive generative models leverage vast global pre-training over international posthumanist and philosophical literature, allowing them to map broad semantic distributions with high precision, whereas smaller encoders like RoBERT require denser thematic patterns to balance abstract, low-frequency classes.

7 Discussion and Qualitative Analysis

7.1 Architectural Error Propagation: Flat vs. Cascading Hybrids

The performance discrepancy between the Flat Hybrid model (F_1: 0.673) and the *Cascading Hybrid* architecture (F_1: 0.600) highlights a fundamental challenge in multi-stage poetry parsing: sequential error propagation. The cascading framework relies on a binary gate to isolate enjambment candidates before invoking the fine-grained 13-class specialist.

Qualitative error audit reveals that contemporary free verse heavily employs ambiguous line transitions that cause hard classification bottlenecks. If the binary layer misclassifies a subtle structural break as `NONE`, the instance is permanently pruned from the pipeline. The downstream multi-class specialist never receives the token sequence, rendering it incapable of recovering the false negative. Conversely, the flat model architecture evaluates all 14 boundaries simultaneously within a unified probability space. Using weighted cross-entropy loss avoids structural bottlenecks, making the flat approach significantly more resilient to the non-linear, high-entropy rhythm of contemporary poetry.

7.2 Linguistic Error Analysis in Enjambment Detection

Granular evaluation shows distinct morphosyntactic blind spots separating the specialized encoders from the frontier generative baseline.

Long-Distance Agreements (`subj_verb`). Generalist large language models frequently fail to preserve clausal dependencies when they are structurally stretched across metrical boundaries. A prominent example occurs in Poem 110, where the subject is separated from its predicate by a dense prepositional modifier sequence spanning multiple lines:

> *simpla mea prezență printre ființele vii ale lumii*
> ['my simple presence among the living beings of the world']
> | *le-ar șterge pe toate de pe suprafața pământului.*
> ['would erase them all from the face of the earth.']

The expert annotation registers this as a clear `subj_verb` enjambment due to the syntactic incomplete tension at the line break. While our hybrid specialized models successfully decode this long-distance relation, Gemini Pro 3.1 erroneously maps the junction to `NONE`. This indicates that prompt-driven generative models lack the strict morphosyntactic grounding required to track nested structural agreements across line breaks, biasing them toward standard prose phrase boundaries.

The Functional Word Overlook (`pb_relword` and `cross_clause`). Gemini systematically struggles with functional boundaries, underperforming on relative-word disruptions (F_1: 0.633 vs. RoBERT's 0.842). In Poem 1, the line ends on the preposition *în* ['in'], splitting it from its governed noun phrase:

> *chiar nu era nimic frumos sau complicat în* |
> ['there really was nothing beautiful or complicated in']
> *sentimentul ăla).* ['that feeling).']

Gemini misclassifies this junction as a phrase-bounded adjective-preposition break (`pb_adj_prep`), mistaking the long-distance adjective *complicat* as the direct governor of the preposition. Similarly, on clausal boundary breaks

(cross_clause), such as in Poem 427 (*odată ajunși în cafenea* ['once arrived in the cafe'] | *hainele lui...* ['his clothes']), Gemini predicts NONE. Because frontier models are highly trained on standard prose where text flows linearly, their inner attention distributions struggle to weight the rhythmic pause when functional or clausal markers are left dangling at the end of a verse.

7.3 Thematic Mapping: Cultural Grounding vs. Global Semantic Spaces

The thematic task showcases a sharp behavioral split between localized domain adaptation and mass parameter scales. The fine-tuned RoBERT model decisively beats Gemini Pro 3.1 on the dominant macro-theme *Love & Intimacy & Mundane Life* (F_1: 0.760 vs. 0.538). This stems from the concept of **cultural grounding**. Post-2000 Romanian poetry frequently anchors emotional intimacy within visceral, localized domestic realities. For instance, in Poem 857, the text shifts between raw, mundane commands (*dă-i peste bot, alege viața confortabilă* ['hit it over the snout, choose the comfortable life']) and identity deconstruction. Expert human annotators assigned the dominant label *Identity & Otherness*. Gemini Pro 3.1, however, misclassifies it as *Philosophical & Ontological Inquiry*. The generative model operates within a global semantic space; it identifies high-level abstract metaphors (e.g., *parabolele* ['the parables']) and defaults to a broader philosophical category, completely missing the stylistic domestic markers that specify post-communist mundane realism.

Conversely, Gemini's massive scale proves superior when tracking contemporary conceptual fields. In Poem 674, a highly condensed text explicitly introduces modern digital tokens alongside spiritual invocations:

> *Trimite-mi, Doamne, un emoticon*
> ['Send me, Lord, an emoticon']
> *cu un înger, sau o inimă, sau măcar un avocado*
> ['with an angel, or a heart, or at least an avocado']

Our specialized RoBERT encoder misclassifies this text into *Identity & Otherness*, getting distracted by the heavy activation of the religious token *Doamne* ['Lord'], which historically correlates with existential themes in Romanian literature. Gemini Pro 3.1, utilizing its extensive multilingual pre-training, perfectly maps the lexical co-occurrence of *emoticon* and *avocado* to identify the dominant theme of *Digital Realities & Posthumanism*. This demonstrates that while specialized encoders excel at capturing localized cultural idioms, large language models maintain an advantage in mapping globalized, technology-driven semantic domains.

7.4 Diachronic and Cross-Lingual Extensibility

While our RoBERT encoder is optimized specifically for contemporary Romanian, the underlying methodology— the structural taxonomy of enjambment and the distant supervision pipeline—is inherently language-agnostic. As observed in recent diachronic studies (Fatima et al., 2026), decoupling broad generational stylistic shifts from localized authorial idiolects requires robust, author-balanced corpora. Our RoPoetry2K dataset strictly adheres to this principle, ensuring that the detected syntactic ruptures reflect broad post-2000 literary movements rather than individual idiosyncrasies.

8 Conclusion

We presented a comprehensive computational framework for the stylistic and semantic analysis of contemporary Romanian poetry, addressing the dual challenges of fine-grained enjambment detection and macro-thematic classification. We introduced a novel, human-annotated corpus capturing the post-2000 literary landscape and demonstrated that smaller, domain-adapted language models powered by programmatic weak supervision can successfully decode high-entropy literary syntax.

Our results challenge the assumption that massive frontier models render specialized architectures obsolete. Instead, we prove that at least on morphosyntactically complex poetic datasets, a mixed approach is strictly necessary. While Gemini Pro 3.1 excels at mapping broad, abstract global semantic fields, our optimized RoBERT pipelines systematically outperform the generalist LLM on complex native syntactic dependencies and localized cultural idioms. By seamlessly ensembling these specialized BERT-like models with generative LLMs, our framework successfully bridges granular syntax and global semantics. Furthermore, by formalizing enjambment detection as a boundary classification task anchored in universal syntactic cohesion, our methodology provides a scalable template. Future applications can easily export this hybrid distant-supervision framework to widely spoken languages, proving that models developed in flexible-syntax testbeds can drive robust solutions for global computational poetics.

Limitations

Despite the robust empirical performance of our proposed methodologies, several limitations must be acknowledged. First, while our corpus and thematic cultural markers are anchored within the contemporary Romanian post-2000 literary landscape, our syntactic enjambment taxonomy is inherently generalizable. Romanian serves as a rigorous testbed due to its highly flexible word order (permitting both noun-adjective and adjective-noun inversions), which maximizes the structural variance of poetic ruptures. The primary language-specific constraint is the Romanian enclitic definite article (e.g., *omul* instead of *the man*), which naturally precludes the article-noun enjambments common in languages with proclitic articles. Consequently, rather than being limited by these morphosyntactic complexities, our core taxonomy establishes a comprehensive framework that can be seamlessly adapted to languages with more rigid syntactic paradigms, such as English.

Second, our programmatic distant supervision pipeline relies on the standard `ro_core_news_lg` spaCy framework. Because standard dependency parsers are inherently trained and optimized on normative prose, the complete omission of punctuation, heavy fragmentation, and syntactic displacements typical of modern free verse can introduce localized parsing errors into the Silver dataset generation. Although our strict `ABSTAIN` pruning strategy successfully minimizes this noise, it comes at the cost of coverage.

Finally, our large language model baseline evaluation was restricted to a zero-shot regime utilizing Gemini Pro 3.1. Due to the proprietary, closed-weights nature of commercial frontier models, invisible downstream API updates or weights shifts executed by the providers over time may affect long-term reproducibility (Chen et al., 2024a). Furthermore, this black-box paradigm introduces a significant risk of data contamination. Because our corpus is sourced from published post-2000 volumes and accessible digital literary archives, it is highly probable that the LLM encountered fractions of these primary texts, or their associated literary critiques, during its massive multilingual pre-training phase. This potential memorization could artificially inflate the LLM's zero-shot thematic mapping capabilities, underscoring the systemic vulnerability of benchmarking specialized tasks against commercial cloud-based models.

Ethical Considerations

The texts collected in the **RoPoetry2K** corpus are subject to authorial copyright. In compliance with ethical NLP research practices and the ACL Code of Ethics, the corpus was compiled strictly for non-commercial, academic text-and-data-mining purposes, falling under fair use provisions for computational research. To respect intellectual property rights, the publicly released version of the dataset will provide the structural annotations alongside strictly necessary text snippets or URL pointers to the original publishers, rather than redistributing full copyrighted volumes. Furthermore, the dataset was manually audited to ensure it does not contain sensitive Personally Identifiable Information (PII) beyond the publicly established and credited names of the authors.

To foster reproducibility and support lesser-resourced linguistic research, the complete **RoPoetry2K** dataset, the *Gold* annotations, and the comprehensive training and evaluation scripts will be made publicly available in an open-source repository upon the acceptance and publication of this manuscript.

AI Assistance Disclosure. We disclose that AI assistance was utilized during the preparation of this manuscript solely for text editing, language translation, formatting compression, and code assistance.

References

Rodrigo Agerri, Josu Bermudez, and German Rigau. 2014. IXA pipeline: Efficient and ready to use multilingual NLP tools. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3823–3828, Reykjavik, Iceland. European Language Resources Association (ELRA).

Amruta Barbadikar and Amba Kulkarni. 2024. Automatic Sanskrit poetry classification based on kāvyaguṇa. In *Proceedings of the 21st International Conference on Natural Language Processing (ICON)*, pages 109–119, AU-KBC Research Centre, Chennai, India. NLP Association of India (NLPAI).

Timo Baumann, Hussein Hussein, and Burkhard Meyer-Sickendiek. 2018. Analysing the focus of a hierarchical attention network: the importance of enjambments when classifying post-modern poetry. pages 2162–2166.

Sriharsh Bhyravajjula, Melanie Walsh, Anna Preus, and Maria Antoniak. 2025. so much depends / upon / a

whitespace: Why whitespace matters for poets and LLMs. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 35156–35173, Suzhou, China. Association for Computational Linguistics.

Lingjiao Chen, Matei Zaharia, and James Zou. 2024a. How Is ChatGPT's Behavior Changing Over Time? *Harvard Data Science Review*, 6(2). Https://hdsr.mitpress.mit.edu/pub/y95zitmz.

Yanran Chen, Hannes Gröner, Sina Zarrieß, and Steffen Eger. 2024b. Evaluating diversity in automatic poetry generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 19671–19692, Miami, Florida, USA. Association for Computational Linguistics.

Alina Maria Ciobanu, Anca Dinu, and Liviu Dinu. 2014. Predicting Romanian stress assignment. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 64–68, Gothenburg, Sweden. Association for Computational Linguistics.

Marius Conkan. 2022. The rhizomatic: A spatial form in contemporary romanian poetry. *Metacritic Journal for Comparative Studies and Theory*, 8:166–187.

Mihai Dinu. 1986. *Ritm și rimă în poezia românească*. Cartea Românească, București.

Stefanie Dipper and Ronja Laarmann-Quante. 2024. UD for German poetry. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 177–188, Miami, USA. Association for Computational Linguistics.

Noor Fatima, Hasan Faraz Khan, and Irfan Ahmad. 2026. From classical to contemporary: Evolutionary analysis & classification of Urdu poetry. In *Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script*, pages 182–191, Rabat, Morocco. Association for Computational Linguistics.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, and 181 others. 2023. Gemini: A Family of Highly Capable Multimodal Models. *arXiv e-prints*, arXiv:2312.11805.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spacy: Industrial-strength natural language processing in python. *Zenodo*.

Jack Hopkins and Douwe Kiela. 2017. Automatically generating rhythmic verse with neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 168–178, Vancouver, Canada. Association for Computational Linguistics.

Chihan Huang and Xiaobo Shen. 2025. PoemBERT: A dynamic masking content and ratio based semantic language model for Chinese poem generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 50–60, Abu Dhabi, UAE. Association for Computational Linguistics.

Nora Iuga. 2019. Nora iuga: „singura definiție a marii poezii este tăcerea". România literară nr. 22/2019. Interviewed by Laurențiu-Ciprian Tudor.

Ruth Koops van' t Jagt, John Hoeks, Gillis Dorleijn, and Petra Hendriks. 2014. Look before you leap: How enjambment affects the reading of poetry. *Scientific Study of Literature*, 4.

Emanuel Lupașcu. 2022. Postumanismul și poezia română contemporană. *Transilvania*, (6-7):44–57.

Solomon Marcus. 1970. *POETICA MATEMATICĂ*. EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA BUCUREŞTI, București.

Mihai Masala, Stefan Ruseti, and Mihai Dascalu. 2020. RoBERT – a Romanian BERT model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6626–6637, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Irene Miani, Sara Stymne, and Gregory R. Darwin. 2025. Cross-genre learning for Old English poetry POS tagging. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 708–724, Vienna, Austria. Association for Computational Linguistics.

Antonio Quilis. 1964. *Estructura Del Encabalgamiento En La Métrica Española; Contribución a Su Estudio Experimental*. Consejo Superior de Investigaciones Científicas,Patronato "Menéndez y Pelayo," Instituto "Miguel de Cervantes,", Madrid.

Pablo Ruiz Fabo, Clara Martínez Cantón, Thierry Poibeau, and Elena González-Blanco. 2017. Enjambment detection in a large diachronic corpus of Spanish sonnets. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 27–32, Vancouver, Canada. Association for Computational Linguistics.

K. Spang. 1983. *Ritmo y versificación. Teoría y práctica del análisis métrico*. Universidad de Murcia, Spain, Spain.

Valentin Tașcu. 2006. *TEORIA RITMURILOR ÎN ARTĂ ŞI POEZIE*. Diacronia, București.

Ping Wang, Shitou Zhang, Zuchao Li, and Jingrui Hou. 2023. Enhancing Ancient Chinese understanding with derived noisy syntax trees. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 83–92, Toronto, Canada. Association for Computational Linguistics.

A Appendix

A.1 Annotator Demographics, Expertise, and Guidelines

To ensure high-quality and reproducible labels for the highly subjective tasks of contemporary poetry analysis, a formal annotation manual was compiled prior to the labeling phase. This manual established strict operational definitions, linguistic constituency tests for each enjambment category, and semantic boundary criteria for the macro-themes.

To comply with responsible NLP research practices, we report the annotator demographics. The evaluation panel consisted of three domain experts: a lead annotator (a published Romanian poet) and two independent philologists, with degrees in linguistics or literature. All three annotators are native Romanian speakers (L1). Because the annotation was conducted by domain experts collaborating within a university research framework, no crowdsourcing platforms or hourly compensation structures were utilized.

Annotation Process and Inter-Rater Agreement (IRA). We adopted a two-phase *Pilot-to-Expert* methodology to validate the guidelines and establish the ground truth.

Phase 1 (Pilot and Calibration): The three experts independently annotated a shared, randomly sampled subset to measure the reliability of the annotation schema and refine the guidelines.

- *Thematic Classification:* The panel initially applied the exploratory 8-class multi-label framework to a subset of 100 poems. Upon empirical observation of high thematic density (averaging 2.95 labels per poem) and subsequent semantic noise, the panel decided to execute the consolidation strategy, mapping the annotations into the final 5 mutually exclusive macro-classes. On this finalized single-label schema, the panel achieved substantial agreement ($\kappa = 0.78$) via Fleiss' Kappa.
- *Enjambment Detection:* On an initial set of 50 fully scanned poems, agreement for the binary presence of enjambment was near-perfect ($\kappa = 0.88$), while the fine-grained 14-class syntactic taxonomy yielded strong agreement ($\kappa = 0.74$).

Following the independent labeling, all disagreements in the pilot phase were systematically extracted and adjudicated through panel consensus, led by the primary expert, to refine the annotation

guidelines. For the thematic task, to successfully transition from the 8-class multi-label schema to the 5-class single-label benchmark, the panel established a hierarchical prioritization rule. This rule dictates that specific, localized thematic markers must consistently supersede broader, general categories when resolving the single dominant label. Similarly, for the enjambment task, the pilot revealed recurrent ambiguities in distinguishing between a verbal modifier and a prepositional regime (`pb_verb_adv` vs. `pb_verb_cprep`) in elliptical free verse. To resolve this, the panel generated a strict substitution-based constituency test.

Phase 2 (Task-Specific Expert Annotation): Following guideline calibration, the main annotation phase was consolidated to the lead domain expert. This architectural decision was taken to guarantee absolute stylistic consistency across the nuanced 14-class syntactic taxonomy and the dense cultural markers of the thematic classes.

- *For Thematic Classification:* The lead expert evaluated the remaining 928 poems, yielding the fully annotated RoPoetry2K.
- *For Enjambment Detection:* The lead expert finalized the Gold Standard, executing targeted sampling on rare syntactic classes to expand the initial pilot set to the final 107-poem dataset (1,537 line pairs).

By anchoring the bulk of the annotation in the adjudicated consensus of the pilot phase, this process ensures that the datasets reflect reproducible linguistic rules rather than an isolated subjective interpretation.

A.2 Detailed Thematic Taxonomy and Multi-Label Statistics

The initial annotation pipeline for the corpus was built upon eight distinct taxonomic themes derived from contemporary analyses of post-2000 verse (Lupașcu, 2022): *Digital Realities & Posthumanism*, *Eco-Consciousness*, *Identity & Otherness*, *Interconnected Structures*, *Philosophical & Ontological Inquiry*, *Mundane & Daily Life*, *Love & Intimacy*, and *Self-Referentiality (Ars Poetica)*.

During the exploratory multi-label phase, human annotators mapped the high structural and thematic density of the texts, leading to an average of 2.95 labels per poem. Preliminary document frequency distributions revealed that wide-scope categories dominated the unpruned dataset: *Philosophical Inquiry* appeared in 67.02% of all poems, followed closely by *Mundane Life* (58.75%) and *Identity* (57.29%). Conversely, specialized categories suffered from extreme low-frequency density, such as *Self-Referentiality* (16.14%) and *Eco-Consciousness* (9.72%), co-occurring almost exclusively within larger semantic frameworks.

This heavy overlap created high linguistic noise and representation sparsity, which motivated the collapsing strategy to safeguard downstream machine learning evaluation from severe gradient instability. The consolidation resolved semantic overlaps by applying the following structural logic:

- *Love & Intimacy* and *Mundane & Daily Life* were merged into a single macro-category (*Love & Intimacy & Mundane Life*), as contemporary Romanian poetry heavily roots emotional experiences within everyday reality (Lupașcu, 2022).
- Low-frequency topics like *Eco-Consciousness* (9.72%) and *Interconnected Structures* (22.95%) were integrated into *Digital Realities & Posthumanism*. Conceptually, modern eco-poetry and the rhizomatic spatial forms identified in recent Romanian verse (Conkan, 2022) are deeply tied to posthumanist anxieties and techno-ecological environments.
- *Identity & Otherness*, *Self-Referentiality*, and *Philosophical Inquiry* were maintained as standalone classes, as their boundaries remained distinct during the annotation process.

A.3 Programmatic Labeling Function Evaluation

For our programmatic weak supervision frameworks, we make a clear distinction between the *rule coverage* and standard *recall* metrics. Coverage denotes the percentage of the dataset where the labeling function (lf_spacy) returns a definitive class rather than the `ABSTAIN` signal. On our Gold Standard of 1,537 line pairs, the function achieved a coverage of 42.55% (654 predicted instances).

Because the primary objective of silver dataset generation is to maintain a high signal-to-noise ratio, the heuristic rules are intentionally conservative. To explicitly illustrate the trade-off between coverage, False Positives (FP), and False Negatives (FN), Table 4 presents the macro-level confusion matrix evaluating the programmatic predictions against the Gold Standard. For clarity, the 13 fine-grained enjambment classes are aggregated into a single *Enjambment* macro-class.

The evaluation reveals a highly specialized noise profile, typical for distant supervision heuris-

True (Gold)	Predicted (lf_spacy)		
	NONE	Enjambment	ABSTAIN
NONE (1089)	245	193	651
Enjambment (448)	2	214	232

Table 4: Macro-level Confusion Matrix for the programmatic labeling function evaluated on the Gold Standard.

tics. The function achieves near-perfect precision (99.1%) when predicting the absence of a structural break (NONE), committing only 2 critical FN (predicting NONE when an actual enjambment occurred).

Conversely, the FP rate is concentrated within the positive classes: 193 instances were incorrectly flagged as enjambments due to the parser misinterpreting standard line-end punctuation or prose-like formatting as severed syntactic ties. The 883 ABSTAIN instances act as a strict safety buffer, filtering out highly ambiguous poetic inversions. Ultimately, this programmatic labeling function provides a massive, structurally sound foundation; the inherent FP introduced by the spaCy parser are subsequently smoothed out by the neural encoder during the weighted cross-entropy optimization.

To provide full transparency regarding the heuristics' behavior across specific syntactic structures, Table 5 details the granular classification report strictly on the 654 instances that bypassed the ABSTAIN filter.

Category	Precision	Recall	F1	Support
NONE	0.99	0.56	0.71	438
lexical	1.00	1.00	1.00	20
pb_adj_adv	0.88	0.88	0.88	8
pb_adj_noun	0.59	1.00	0.74	10
pb_adj_prep	0.18	0.37	0.24	19
pb_adv_prep	0.59	1.00	0.74	13
pb_noun_prep	0.22	0.91	0.36	22
pb_verb_adv	0.57	0.57	0.57	7
pb_verb_chain	0.89	1.00	0.94	17
pb_verb_cprep	0.22	0.73	0.34	11
pb_relword	0.89	0.77	0.83	31
cross_clause	0.30	0.78	0.44	32
subj_verb	0.43	0.82	0.56	11
obj_verb	0.67	0.93	0.78	15
Accuracy		0.65		654
Macro Avg	0.60	0.81	0.65	654
Weighted Avg	0.85	0.65	0.69	654

Table 5: Granular performance report of the programmatic labeling function (lf_spacy) on the Gold Standard, calculated exclusively on the active coverage zone (654 instances).

The granular metrics confirm that while explicit structural elements like hyphenated breaks (lexical) or short relative pronouns (pb_relword) are captured with high precision (1.00 and 0.89, respectively), rule-based parsing severely struggles with prepositional regimes (e.g., pb_adj_prep at 0.18 precision). This sharp drop in minority-class precision empirically validates our architectural decision: a purely rule-based system cannot decode the fluid syntax of modern verse, necessitating the integration of a neural encoder (RoBERT) capable of smoothing out these localized False Positives through contextual embedding representations.

A.4 Enjambment Taxonomy and Annotated Examples

Table 6 presents the distribution of enjambment categories in the RoPoetry2K corpus:

Category	Count	%
NONE	1,089	70.85%
lexical	21	1.37%
pb_adj_adv	15	0.98%
pb_adj_noun	29	1.89%
pb_adj_prep	20	1.30%
pb_adv_prep	13	0.85%
pb_noun_prep	26	1.69%
pb_verb_adv	20	1.30%
pb_verb_chain	21	1.37%
pb_verb_cprep	20	1.30%
pb_relword	82	5.34%
cross_clause	84	5.47%
subj_verb	47	3.06%
obj_verb	50	3.25%

Table 6: Distribution of enjambment categories in the dataset.

Table 7 combines the definitive linguistic boundaries of our 14-class taxonomy with validated examples harvested from the contemporary corpus.

A.5 Corpus Sources and Metadata

This appendix section provides the complete list of primary sources used to compile the contemporary Romanian poetry corpus.

Digital Portals and Web Sources. Table 8 lists the digital platforms, literary magazines, and online archives from which poems were collected. All web resources were accessed between January and February 2026.

Physical Books and Anthologies Table 9 details the physical books, single-author volumes, and collective anthologies from which texts were manually transcribed or digitized.

Label	Enjambment form	Example	Poem ID
lexical	A rupture inside the same word.	între două diete [spar- | tane] *[between two [Spar- | tan] diets]*	950
pb_adj_adv	A rupture between a modifying adverb and the adjective it qualifies.	foarte tânără și [foarte$_{adv}$ | dornică$_{adj}$] de afirmare. *[very young and [very$_{adv}$ | eager$_{adj}$] to assert herself]*	264
pb_adj_noun	The separation of a noun from its modifying adjective or determiner, regardless of whether the adjective precedes or follows the noun.	ca un [maratonist$_{noun}$ | Neobosit$_{adj}$] *[like a [tireless$_{adj}$ | marathon$_{noun}$] runner]*	820
pb_adj_prep	A break between an adjective (or a past participle functioning as an adjective) and the prepositional phrase that complements it.	nu le-a mers bine decât [inundate$_{adj}$ | de$_{prep}$] lumina romantică a celei de-a doua căsnicii *[things never went well for them except when [flooded$_{adj}$ | by$_{prep}$] the romantic light of the second marriage]*	795
pb_adv_prep	The separation of an adverb from its governing prepositional complement.	să clipesc atât de [aproape$_{adv}$ | de$_{prep}$] obrazul tău *[to blink [so close$_{adv}$ | to$_{prep}$] your cheek]*	7
pb_noun_prep	A rupture between a noun and the prepositional phrase acting as its attribute.	Noul [Ierusalim$_{noun}$ | cu$_{prep}$] rinoceri și lei *[The New [Jerusalem$_{noun}$ | with$_{prep}$] rhinoceroses and lions]*	620
pb_verb_adv	A break between a verb and a direct modifying adverb (adverbial modifier).	după fumul și fluierul | ce se vor [auzi$_{verb}$ | mereu$_{adv}$] *[by the smoke and the whistle | that will [always$_{adv}$ | be$_{verb}$] heard]*	100
pb_verb_chain	The separation of an auxiliary verb from the main lexical verb in a compound tense or verbal chain.	atunci ți-aș spune că pereții s-[au$_{aux}$ | fisurat$_{verb}$] cu încetinitorul *[then I'd tell you that the walls [were$_{aux}$ | splitting$_{verb}$] in slow motion]*	1
pb_verb_cprep	A rupture between a verb and the specific preposition required by its valency (prepositional regime).	(Cel mai mult mă agasează că nu au de-a [face$_{verb}$ | cu$_{prep}$] mine) *[(Most of all, it bothers me that they have nothing to [do$_{verb}$ | with$_{prep}$] me)]*	1
pb_relword	A break between a functional word (conjunction, particle, or short clitic pronoun) and the lexical word it is syntactically attached to.	Dacă cineva ar încerca să-[mi$_{pron}$ | povestească$_{verb}$] viața pe Wikipedia *[If someone tried to [tell$_{verb}$ | my$_{pron}$] life story on Wikipedia]* a trăi e mai important [decât$_{con}$ | a$_{aux}$] ști pentru ce trăiești *[Living is more important [than $_{con}$ | knowing$_{verb}$] what you live for]*	56 264
cross_clause	A break between two clauses, with the coordinating or subordinating conjunction introducing the second line.	refuz să-mi ascult [doctorul$_{noun}$ | care$_{rel}$]-a insistat să vizitez *[I refuse to listen to my [doctor$_{noun}$ | who$_{rel}$] insisted I visit]*	78
subj_verb	The separation of the subject from the predicate verb, despite the distance between them.	[dragostea$_{subj}$ | coboară$_{verb}$] încet, cu precauție, ca un alpinist *[[love$_{subj}$ | descends$_{verb}$] slowly, with caution, like an alpinist]*	40
obj_verb	The separation of the direct object from its governing verb, despite the distance between them.	Am [iubit$_{verb}$ | un singur bărbat$_{obj}$] *[I [loved$_{verb}$ | just one man$_{obj}$]]*	56
NONE	There is no stylistic rupture between the two verses.	ce se preface că nu știe nimic | despre dezastrul geopolitic în care suntem captivi *[who pretends to know nothing | about the geopolitical disaster in which we are trapped]*	78

Table 7: Comprehensive taxonomy of enjambment types with linguistic definitions and illustrative examples.

Platform Name	Primary URL
O Mie de Semne	https://omiedesemne.ro/
Vatra Magazine	https://revistavatra.org
Echinox Magazine	https://revistaechinox.ro/
Familia Magazine	https://revistafamilia.ro/
Scena9	https://www.scena9.ro/
Bookaholic	https://www.bookaholic.ro/

Table 8: Digital platforms used for text collection.

A.6 Implementation Details and Hyperparameters

Our specialized models are built upon the `readerbench/RoBERT-base` checkpoint, which contains approximately 110 million parameters, consisting of 12 attention layers, 768 hidden dimensions, and 12 attention heads. All fine-tuning and distant supervision experiments were executed locally on an Apple MacBook Pro equipped with an M3 Pro chip and 18GB of unified memory, utilizing Metal Performance Shaders (MPS) for hardware-accelerated training. Due to the targeted architectural design and the efficiency of the MPS backend, the computational overhead is minimal; a complete 5-fold cross-validation loop for enjambment detection requires approximately 1 hour of compute time.

To handle class imbalances during fine-tuning, the custom Weighted Cross-Entropy Loss computes the penalty weights W_c for each class as follows:

$$W_c = \frac{N_{\text{total}}}{C \times N_c} \tag{1}$$

where C is the total number of classes in the active taxonomy, N_{total} represents the total instance count, and N_c is the document frequency of class c.

Across all architectural variants, network optimization is driven by the **AdamW** optimizer with a base learning rate of $2e-5$ and a weight decay of 0.01. Training performance is monitored via the validation Macro F_1-score at the end of each epoch, triggering an automated rollback to the best historical weights upon completion. To safeguard against overfitting, an Early Stopping callback is enforced, terminating execution if the validation Macro F_1-score fails to improve for 3 consecutive epochs.

Task-Specific Configurations. For Task 1 (Enjambment Detection), the sequence length for line-pair junctions is capped at 128 tokens. Minority categories containing fewer than 15 examples are subject to automated data oversampling prior to fold execution. The models are trained with a batch size of 16 for a maximum of 15 epochs per fold.

For Task 2 (Thematic Classification), the sliding window text-chunking strategy utilizes a maximum sequence length of 512 tokens and a stride (overlap) of 128 tokens. The network is fine-tuned for a maximum of 12 epochs with a batch size of 8, employing a cosine learning rate scheduler with a warmup ratio of 0.1. During inference, the post-processing prediction multipliers applied to the max-pooled probabilities are optimized via grid-search, scanning intervals between 0.8 and 1.8 at 0.1 increments mapped strictly against the validation set.

A.7 Gemini Pro 3.1 Prompt Templates

To ensure the complete reproducibility of our study, this section details the prompt templates utilized for the zero-shot LLM baselines. Because the target corpus consists entirely of contemporary Romanian verse, the system prompts were executed natively in Romanian to optimize the model's morphosyntactic alignment. Below, we provide both the original execution text and an English translation.

A.7.1 Task 1: Fine-Grained Enjambment Detection Prompt

Original Romanian Execution Prompt:

```
Rol: Ești un expert în filologie
română și procesarea limbajului
natural, specializat în analiza poeziei
contemporane.


Task: Analizează fragmentul de poem de
mai jos și identifică toate instanțele
de ingambament. Ingambamentul apare
atunci când o unitate sintactică strânsă
este ruptă de sfârșitul versului.


Tipologie de utilizat :
0. NONE: Nu există ingambament între
versul N și N+1. De exemplu: "mi se
rupe inima | doamne dumnezeule".
1. lexical: Cuvânt tăiat de cratimă,
dar nu neapărat, la final de rând. De
exemplu: "pola- | rizați".
2. pb_adj_adv: Adverb + Adjectiv,
Adverb intensificator. Exemplu: "foarte
| frumos".
3. pb_adj_noun: Substantiv + Adjectiv.
Exemplu: "lumea | ei" sau "lumea |
întreagă".
4. pb_adj_prep: Adjectiv/Participiu +
Prepoziție (Atenție: dacă participiul
descrie un substantiv, e pb_adj_prep,
nu verb!). Exemplu: "plină | de".
5. pb_adv_prep: Adverb + Prepoziție.
Exemplu: "departe | de".
6. pb_noun_prep: Substantiv +
Prepoziția sa atributivă. Exemplu: "ora
```

Title	Author(s) / Editor(s)	Publisher	Year
HABITAT	Aleksandar Stoicovici	ART	2025
ÎNSEMNĂRI PE TIVUL FUSTEI	Anca-Ioana Câdă	ART	2025
PĂMÂNTUL PLATH	Anca Zaharia	ART	2025
scrisoare de restituire	Andreea Catrina	Cartex	2025
PANOU	Andrei Doboș	ART	2025
filme coreene despre răzbunare	Andrei Petrea	Cartex	2025
VALEA SĂPUNULUI	Andrei Zbîrnea	ART	2025
HATE ME LOVE	Cosmin Perța	ART	2025
singurătatea vine pe facebook	Dan Mircea Cipariu	Tracus Arte	2012
NU VA RĂMÂNE NIMIC	Liviu Diamandi, Dmitri Miticov	ART	2025
Nu închide geamul	FELICIA SIMION	Cartex	2025
MARUTH	Marius Aldea	ART	2025
ATÉ (O EDUCAȚIE)	Olga Ștefan	ART	2025
PREDETERMINAT	Teodora Coman	ART	2025
Liternauții	Ioana Nicolaie	Paralela 45	2025
1.20.2024. Contemporary Romanian Poetry	Carmen Corbu, Corina Taraş-Lungu	Fundația Culturală „Augustin Buzura”	2024
când va veni ceea ce este desăvârşit	Andrei Dósa	Tracus Arte	2011

Table 9: Physical books and print anthologies included in the corpus.

| de”.
7. pb_verb_adv: Verb + Adverb. Exemplu: ”alerg | repede”.
8. pb_verb_chain: Auxiliar + Verb. Exemplu: ”am | terminat”.
9. pb_verb_cprep: Verb + Prepoziția regimului său verbal. Exemplu: ”de-a face | cu”.
10. pb_relword: Conjuncție (că, să, și, dacă), Particulă (ca, decât) sau Pronume scurt (mă, ne, îmi) la final de vers.
11. cross_clause: Versul se termină înainte de conjuncție (Exemplu: ”Oamenii au decis să înregistreze un videoclip | în care vorbesc despre ei”). Aici ruptura este între cele două propoziții.
12. subj_verb: Subiect + Verb. Exemplu: ”eu | merg”, dar și ”Dar nu știau că cei care vizionează videoclipul | nu sunt prietenoși.”.
13. obj_verb: Obiect + Verb. Exemplu: ”au simțit | pericolul”, dar și ”și buzele au pe dinăuntru | acele striații necesare.”.

Instrucțiuni de Gândire (CoT):
Pentru fiecare joncțiune între versul N și N+1:
1. Extrage ultimele cuvinte din versul N și primele din versul N+1.
2. Verifică dacă există o ”tensiune de completare”. Dacă citești doar versul N, simți nevoia gramaticală de a continua imediat?
3. Identifică relația sintactică. Chiar dacă parserul ar eșua, bazează-te pe logica limbii române.

CONSTRÂNGERI DE OUTPUT (STRICT):
Răspunde EXCLUSIV sub formă de obiect JSON.
Nu adăuga nicio introducere, nicio explicație și niciun comentariu după JSON.
Păstrează structura originală, completând doar câmpul ”tip”.
Folosește DOAR numele exacte ale etichetelor din tipologia de mai sus (ex: pb_adj_adv, nu „Adverb-Adjectiv”).

DATE DE ANALIZAT:

English Translation:

Role: You are an expert in Romanian philology and natural language processing, specialized in the analysis of contemporary poetry.

Task: Analyze the poem fragment below and identify all instances of enjambment. Enjambment occurs when a closely knit syntactic unit is broken by the end of a line.

Typology to use:
0. NONE: No enjambment between line N and N+1. For example: ”mi se rupe inima | doamne dumnezeule”.
1. lexical: A word cut by a hyphen, but not necessarily, at the end of a line. For example: ”pola- | rizați”.
2. pb_adj_adv: Adverb + Adjective, intensifying adverb. Example: ”foarte | frumos”.
3. pb_adj_noun: Noun + Adjective. Example: ”lumea | ei” or ”lumea | întreagă”.
4. pb_adj_prep: Adjective/Participle + Preposition (Note: if the participle describes a noun, it is pb_adj_prep, not a verb!). Example: ”plină | de”.
5. pb_adv_prep: Adverb + Preposition. Example: ”foarte | departe”.
6. pb_noun_prep: Noun + its attributive preposition. Example: ”ora | de”.

7. pb_verb_adv: Verb + Adverb. Example: "alerg | repede".
8. pb_verb_chain: Auxiliary + Verb. Example: "am | terminat".
9. pb_verb_cprep: Verb + the preposition of its verbal regime. Example: "de-a face | cu".
10. pb_relword: Conjunction (că, să, și, dacă), Particle (ca, decât), or Short pronoun (mă, ne, îmi) at the end of a line.
11. cross_clause: The line ends before a conjunction (Example: "Oamenii au decis să înregistreze un videoclip | în care vorbesc despre ei"). Here the break is between the two clauses.
12. subj_verb: Subject + Verb. Example: "eu | merg", but also "Dar nu știau că cei care vizionează videoclipul | nu sunt prietenoși.".
13. obj_verb: Object + Verb. Example: "au simțit | pericolul", but also "și buzele au pe dinăuntru | acele striații necesare.".

Chain of Thought (CoT) Instructions:
For each junction between line N and N+1:
1. Extract the last words from line N and the first words from line N+1.
2. Check if there is a "completion tension". If you read only line N, do you feel a grammatical need to continue immediately?
3. Identify the syntactic relationship. Even if the parser fails, rely on the logic of the Romanian language.

OUTPUT CONSTRAINTS (STRICT):
Respond EXCLUSIVELY in the form of a JSON object.
Do not add any introduction, explanation, or comment after the JSON.
Keep the original structure, filling in only the "tip" field.
Use ONLY the exact label names from the typology above (e.g., pb_adj_adv, not "Adverb-Adjective").

DATA TO ANALYZE:

A.7.2 Task 2: Thematic Classification Prompt

Original Romanian Execution Prompt:

Ești un expert în critică literară.
Clasifică următoarele poezii în una dintre aceste 5 categorii:
- Love & Intimacy & Mundane Life
- Digital Realities & Posthumanism
- Identity & Otherness
- Self-Referentiality (Ars Poetica)
- Philosophical & Ontological Inquiry
Răspunde strict sub formă de tabel JSON cu ID-ul poeziei și tema aleasă.

English Translation:

You are an expert in literary criticism.
Classify the following poems into one of these 5 categories:
- Love & Intimacy & Mundane Life
- Digital Realities & Posthumanism
- Identity & Otherness
- Self-Referentiality (Ars Poetica)
- Philosophical & Ontological Inquiry
Respond strictly in the form of a JSON table with the poem ID and the chosen theme.

A.8 Ensemble Routing Strategy

As discussed in Section 6.1, the *Ensemble* employs a deterministic, rule-based routing heuristic to leverage the complementary architectural strengths of our specialized RoBERT encoder and Gemini Pro 3.1. The routing logic is defined by the architectural strengths of the models: to maximize accuracy, the system defaults to the global semantic mapping capabilities of Gemini Pro 3.1, unless the fine-tuned RoBERT model predicts one of the following eight categories, which represent strict functional and syntactic dependencies:

- **Functional and Clause Boundaries:** cross_clause, pb_relword
- **Prepositional Regimes and Modifiers:** pb_adj_prep, pb_adv_prep, pb_noun_prep, pb_verb_cprep
- **Verbal Adverbs:** pb_verb_adv
- **Negative Baseline (Absence of Enjambment):** NONE

By explicitly delegating these fine-grained structural disruptions to the targeted encoder, the ensemble systematically overrides the LLM's tendency to merge functional poetic line breaks into standard prose boundaries, thereby achieving the highest overall F_1-score.