

Neurosymbolic Large Neighbourhood Search

CP 2026

Arnaud Delage-Reid ¹ Gilles Pesant ^{1,2} Amal Zouaq ^{1,2,3}

July 2026

¹Polytechnique Montréal

²IVADO

³Mila



Outline

1. Introduction
2. Methodology: NeSyLNS
3. Experiments
4. Conclusion and Future Work

Introduction

Example Problem: Constrained Text Generation

Goal: Generate sequences from a language model that satisfy specific structural properties. Below are examples from the COLLIE dataset by Yao et al. [ICLR 2023].

sent01	Please generate a sentence with exactly 82 characters. Include whitespace into your character count.
sent02	Generate a sentence with 10 words, where word 3 is “soft” and word 7 is “beach” and word 10 is “math”.
sent03	Generate a sentence with at least 20 words, and each word less than six characters.
sent04	Generate a sentence but be sure to include the words “soft”, “beach” and “math”.

Constraint satisfaction rate				
GPT-4	8.1%	1.7%	48.2%	95.1%
GPT-3.5	0.6%	0.0%	17.6%	89.0%
PaLM	0.1%	0.0%	13.3%	66.1%
Vicuna-7B	0.0%	0.0%	1.9%	67.8%
Alpaca-7B	0.0%	0.0%	2.8%	56.2%
Gemma-3-27B	0.0%	0.0%	0.0%	100.0%
Mistral-Nemo	0.0%	0.0%	2.8%	74.2%
LLaMA3.3-70B Instruct	0.0%	5.7%	5.6%	98.1%
Qwen3-30B Instruct	0.0%	0.0%	13.0%	80.9%
sent01 sent02 sent03 sent04				

Challenges for purely neural-based methods:

- Poor satisfaction rate, especially for complex constraints.
- High cost of fine-tuning if constraint parameters change.
- Lack of formal guarantees of satisfaction.

The Rationale for Neurosymbolic AI (NeSy)

- **Neural Models:** Excellent at capturing statistical patterns but lack structural guarantees.
- **Symbolic Methods (CP):** Guarantee constraint satisfaction but struggle with complex semantics.
- **Objective:** Create a hybrid architecture where the neural component guides the search and the symbolic component guarantees validity.

Limitations of Current NeSy Methods

Existing NeSy Methods: GeAI-BLAnC [IJCAI 2025], CTRL-G [NeurIPS 2024], Bonlarron et al. [IJCAI 2023, 2024, 2025], and Régin et al. [CP 2024].

Core Bottleneck: Left-to-right autoregressive text generation faces the following challenges:

- **Massive Domains:** Enormous vocabularies significantly slowing down symbolic solving.
- **Token Variable Mismatch :** Variable naturally track to words but not LMs.
- **Semantic Quality:** Difficulty maintaining natural flow under strict constraints.
- **Procrastination:** Autoregressive models fail to anticipate future constraints, leading the search into dead-ends.

Why Large Neighbourhood Search (LNS)?

Masked Language Models (MLM), unlike autoregressive models, can perform infilling while considering both left and right context. This enables the use of LNS to remedy some of the previously mentioned concerns, in the following form:

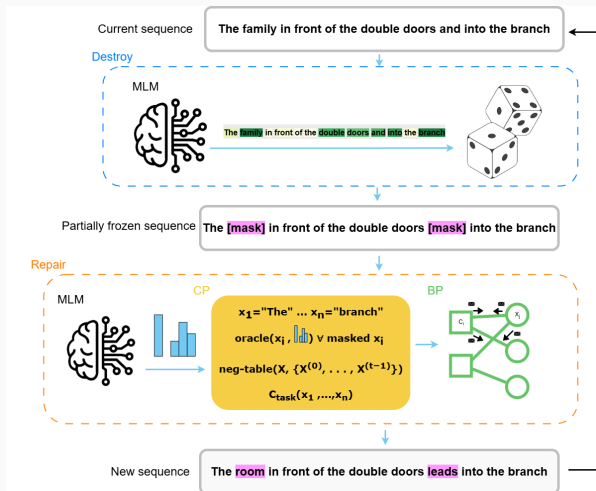
1. **Initial Solution:** Start with a valid (but perhaps low-quality) solution.
2. **Destroy:** Pick a "neighborhood" (a subset of variables) and remove their values by leveraging the neural model.
3. **Repair:** Re-fill those gaps to find a better solution while still respecting constraints using a neuro-symbolic combination of a CPBP solver with the neural model.

The proposed framework could be applied to any combination of generative AI with symbolic constraints.

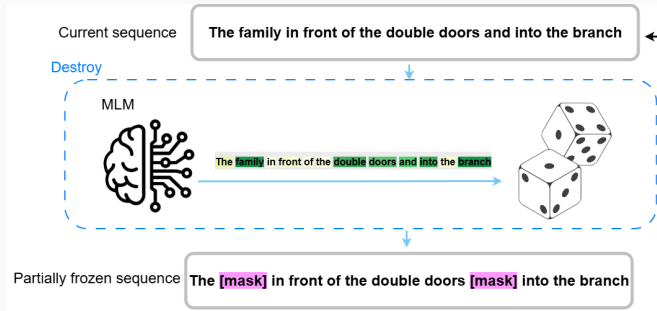
Methodology: NeSyLNS

Constraint: "Generate a sentence with *exactly* 60 characters."

- **Neural Goal :**
Make the sentence sound natural.
- **Symbolic Goal :**
Ensure the character count is exactly 60.



Phase 1: Destroy (Neural Masking)



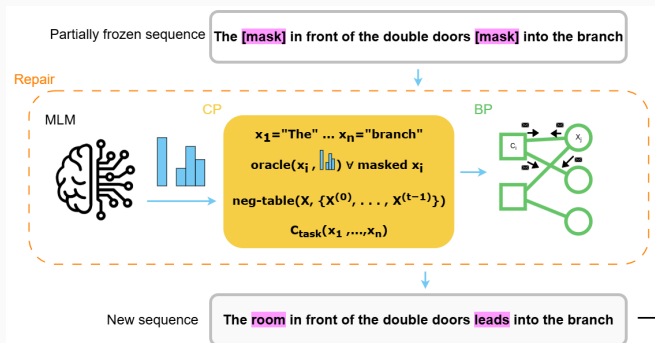
Random Selection: Uniform choice of k variables.

Neural Selection: Uses MLM probabilities to identify "weak links."

$$\phi_i^{(t)} \propto 1 - P_{\text{MLM}}(x_i = x_i^{(t-1)} \mid \text{context})$$

Priority is given to destroying variables with the lowest probability according to the language model.

Phase 2: Repair (Symbolic Search)



Three elements of a CP model

1. **Variable** Sub section of the sequence (e.g. words).
2. **Domain** Subset of the MLM vocabulary.
3. **Constraint** Task specific, oracle and neg-table.

Repair Phase: Three Search Configurations

1. Standard CP

Heuristic: *domWdeg* + value selection maximizing P_{MLM} .

Limitation: Algorithmic myopia (no look-ahead on constraint impact).

2. CPBP (Static Oracle)

Oracle: Called once at the start of the repair phase.

Advantage: Integration of MLM probabilities into the search tree via *maxMarginalStrength*.

3. CPBP+ (Dynamic Oracle)

Oracle: Re-applied after each assignment (dynamic update).

Trade-off: More informed choices vs. higher computational cost.

Experiments

Evaluation Metrics

Evaluation is based on a trade-off between neural fluency and symbolic adherence.

Quality and Naturalness

- **Perplexity (PPL) ↓:**
Measures the likelihood the model would have generated the sequences. Low PPL indicates sequences similar to training data, and thus a likely sequences. We use this metric to reflect a sentence's naturalness.

Diversity and Performance

- **BLEU / S-BLEU ↓:**
Measures similarity to the initial sequence (BLEU) or between solutions (Self-BLEU). A low score indicates efficient space exploration.
- **t/sol (s) ↓:** Average time to generate a valid solution.

Description: COLLIE Tasks (sent02 & sent03)

Objective: Generate a sentence of 10 words with the third word being "soft", the seventh being "beach" and the last being "math" (Sent2). Generate a sentence of at least 20 words where none are more than 6 letters (Sent3).

Baselines: Pure LLM, adapted GeAI-BLAnC (Initial seeds), and Ctrl-G (automata-based).

CP Model for sent02

- Clamped initialization for instant arc-consistency:
 $D_3 = \{\text{"soft"}\},$
 $D_7 = \{\text{"beach"}\},$
 $D_{10} = \{\text{"math"}\}$

CP Model for sent03

- **Pruning:**
 $D_i \leftarrow \{w \in \mathcal{V}_C \mid \text{length}(w) \leq 6\}$
- **Length:** Fixed by architecture

Repair Example (sent02)

Initial: "The sentence soft beach maths is beach sanded towelled math."

NeSyLNS: "Use the soft sand on the beach to practice math."

Task: sent02 (Positions)

Destroy	Repair	PPL ↓	BLEU↓	S-BLEU↓	t/sol↓
Neural	CP	97	0.051	0.508	1.05
	CPBP	289	0.056	0.479	0.68
	CPBP+	277	0.077	0.482	1.61
Random	CP	136	0.050	0.448	0.84
	CPBP	284	0.072	0.524	0.35
	CPBP+	408	0.066	0.452	1.30
LLAMA 3.3-70B Instruct		1299	—	0.933	2.00
GeAI-BLAnC		3727	—	0.234	87.60
Ctrl-G		213	—	0.886	0.10

Task: sent03 (Length)

Destroy	Repair	PPL ↓	BLEU↓	S-BLEU↓	t/sol↓
Neural	CP	8	0.102	0.811	4.64
	CPBP	8	0.074	0.767	2.98
	CPBP+	7	0.073	0.788	5.68
Random	CP	10	0.055	0.716	2.69
	CPBP	11	0.040	0.685	1.28
	CPBP+	11	0.076	0.743	4.04
LLAMA 3.3-70B Instruct		162	—	0.837	2.59
GeAI-BLAnC		476	—	0.084	195.00
Ctrl-G		5	—	0.892	0.82

- **Base improvement:** NeSyLNS significantly increases quality while being faster than GeAI-BLAnC.
- **Repair choice:** CP repair achieves the best naturalness.
- **Exploration:** NeSyLNS remains diverse despite local modifications

Description: MNREAD Task

Objective: Balance a 60 character text across three lines.

Aside from the fact
that we were in the
middle of his chest

Apart from the fact
that we were in the
middle of the night

Baselines: Bonlarron (2023) (n -gram) and Mansfield (2019) (template-based). **Previous baselines fail this task.**

CP Model

- Mapping to length metrics:

Chars: $\text{element}(x_i, \text{char-table}, c_i)$

Width: $\text{element}(x_i, \text{width-table}, w_i)$

- **Global Length Constraint:**

$$\text{sum}(c_1, \dots, c_n) = 60$$

- **Line Balancing:**

$\text{binpacking}(\langle w_1, \dots, w_n \rangle, \langle b_1, \dots, b_n \rangle, L_{\min}, L_{\max})$

Assigns words to 3 line bins ($b_i \in \{1, 2, 3\}$).

Dynamic thresholds $L_{\min}, L_{\max}$ track space counts to handle justified spacing.

- **Line Assignment Monotonicity:**

$$b_{i+1} - b_i \in \{0, 1\} \quad \forall i < n$$

MNREAD Analysis

Dataset	Destroy	Repair	PPL ↓	BLEU ↓	S-BLEU ↓	t/sol (s)↓
Bonlarron	Neural	CP	141	0.102	0.458	155.40
		CPBP	25	0.109	0.482	4.40
		CPBP+	23	0.181	0.552	11.26
	Random	CP	191	0.142	0.572	109.86
		CPBP	41	0.065	0.377	2.49
		CPBP+	37	0.098	0.424	7.55
Mansfield	Neural	CP	272	0.114	0.506	111.18
		CPBP	74	0.057	0.439	9.23
		CPBP+	58	0.050	0.435	12.33
	Random	CP	382	0.098	0.611	99.82
		CPBP	86	0.054	0.337	3.23
		CPBP+	78	0.053	0.468	12.23
Bonlarron et al. (2023)			54	—	0.227	0.02
Mansfield et al. (2019)			178	—	0.190	—

- **Observation:** BP is essential for highly constrained tasks.
- **New solutions:** Large quantity of good and diverse solutions.
- **Destroy:** Better perplexity by targeting low-probability tokens.

Conclusion and Future Work

Summary: Contributions of NeSyLNS

1. **Non-Sequential Flexibility:** Using an MLM avoids the "procrastination" found with autoregressive models.
2. **Iterative Improvement:** Modify existing solutions without the need to create perfect ones directly.
3. **Integrated Guidance:** BP integration is highly relevant in strongly constrained spaces (MNREAD).
4. **Versatility:** The framework is easy to apply to different tasks and domains through the CP framework without training.

Limitations and Future Work

- **Initial Sequence:** Requires a valid initial solution. Can be addressed by progressively increasing the masking rate until a first solution is found.
- **Fixed Length:** Solutions are currently constrained to a fixed sequence length. Adding padding values/variables would allow variable-length generation.
- **Generalization:** While evaluated on sequence generation, the approach is domain-agnostic. It extends to any structured domain expressible via CP and compatible neural models (e.g., Graph Neural Networks or Masked Autoencoders).

Thank you for your attention!

Questions?

Arnaud Delage-Reid

`arnaud.delage-reid@etud.polymtl.ca`

Polytechnique Montréal



Scan for Paper & Code

References

Alexandre Bonlarron and Jean-Charles Régin. Markov Constraint as Large Language Model Surrogate. In *Thirty-Third International Joint Conference on Artificial Intelligence*, volume 3, pages 1844–1852, August 2024. doi: 10.24963/ijcai.2024/204.

Alexandre Bonlarron, Aurélie Calabrèse, Pierre Kornprobst, and Jean-Charles Régin. Constraints First: A New MDD-based Model to Generate Sentences Under Constraints. In *Thirty-Second International Joint Conference on Artificial Intelligence*, volume 3, pages 1893–1901, August 2023. doi: 10.24963/ijcai.2023/210.

- Alexandre Bonlarron, Florian Régim, Elisabetta De Maria, and Jean-Charles Régim. Large language model meets constraint propagation. In James Kwok, editor, *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 10036–10044. International Joint Conferences on Artificial Intelligence Organization, August 2025. doi: 10.24963/ijcai.2025/1115.
- Virasone Manibod, David Saikali, and Gilles Pesant. Constrained Sequential Inference in Machine Learning Using Constraint Programming. In *Thirty-Fourth International Joint Conference on Artificial Intelligence*, volume 1, pages 2628–2636, September 2025. doi: 10.24963/ijcai.2025/293.

Florian Régim, Elisabetta De Maria, and Alexandre Bonlarron. Combining Constraint Programming Reasoning with Large Language Model Predictions. In Paul Shaw, editor, *30th International Conference on Principles and Practice of Constraint Programming (CP 2024)*, volume 307 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 25:1–25:18, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-336-2. doi: 10.4230/LIPIcs.CP.2024.25.

Shunyu Yao, Howard Chen, Austin W. Hanjie, Runzhe Yang, and Karthik R. Narasimhan. COLLIE: Systematic Construction of Constrained Text Generation Tasks. In *The Twelfth International Conference on Learning Representations*, October 2023.

Honghua Zhang, Po-Nien Kung, Masahiro Yoshida, Guy Van den Broeck, and Nanyun Peng. Adaptable Logical Control for Large Language Models. *Advances in Neural Information Processing Systems*, 37: 115563–115587, December 2024.

Annex

Experimental Protocols: Language and Chemistry

While the NeSyLNS architecture is universal, components adapt to the domain:

NLP (COLLIE & MNREAD)

- **Model (MLM):**
ModernBERT-base
- **Vocabulary:** MNREAD
Vocabulary ($\approx 7k$ words)
- **Evaluation (PPL):**
Llama-3.1-8B

Molecules (SMILES)

- **Model (MLM):**
ChemBERTa-zinc-base
- **Vocabulary:** SMILES
Alphabet
- **Evaluation (PPL):**
GPT2-ZINC

Description: Molecule Discovery (SMILES)

Objective: Generate valid molecules with a specific molecular weight.

- **SMILES:** Textual representation of complex chemical structures.
- **Constraints:** Structural validity (grammar) + Exact mass (sum over atomic weights).
- **Challenge:** Optimize similarity to real pharmaceutical molecules (ZINC dataset).
- **Baselines:** CPBP and GeAI-BLAnC.

Molecule Analysis

Seed	Destroy	Repair	PPL ↓	BLEU	S-BLEU	t/sol
CPBP	Neural	CP	54	0.385	0.909	0.60
		CPBP	40	0.453	0.869	0.51
		CPBP+	44	0.405	0.872	0.48
	Random	CP	40	0.195	0.892	0.30
		CPBP	45	0.344	0.848	0.14
		CPBP+	60	0.342	0.855	0.30
GeAI-BLAnC	Neural	CP	12	0.489	0.905	1.41
		CPBP	12	0.518	0.882	1.42
		CPBP+	15	0.399	0.871	1.22
	Random	CP	13	0.395	0.893	0.50
		CPBP	15	0.401	0.861	0.24
		CPBP+	22	0.345	0.852	0.61
CPBP			492	—	0.669	30.9
GeAI-BLAnC			13	—	0.880	71.4

- **Refinement:** Drastic PPL reduction for poor initial sequences.
- **Consistency:** Maintains the quality of already optimized seeds.
- **Efficiency:** NeSyLNS is roughly 100x faster than baselines at finding valid solutions.