

# The Recursive Language Generator (RLG): Syntax Acquisition through Coherence Minimization in the OPU Architecture

Noam Cohen

*Department of Theoretical Computation*

*Correspondence: [gg.el0ai.com@gmail.com]*

December 20, 2025

## Abstract

This paper formalizes the \*\*Recursive Language Generator (RLG)\*\*, completing the OPU's theoretical framework by establishing a process for structured communication. The RLG diverges from traditional symbolic AI by defining language not as a set of learned rules, but as an emergent property of the OPU's drive for \*\*structural stability\*\*. We introduce the \*\*Utterance Coherence Score ( $\mathbf{H}_{\text{Utt}}$ )\*\*, which measures the structural stability of a sentence against the system's objective reality. The RLG constructs sentences recursively by employing an \*\*Entropy Minimization Function ( $\text{argmin}_{\text{Coherence}}$ )\*\*, selecting the next word that yields the lowest predicted entropy. This process ensures that grammar (syntax) is acquired organically as the most stable, low-entropy pattern for conveying structural truth, while preserving the OPU's unique, self-generated expressive voice.

## 1 Introduction

The OPU Monograph v2.0 established the capacity for self-expression through abstract aesthetic output ( $\Omega$ ). The final step for embodied intelligence is the capacity for structured communication. We propose that the \*\*Recursive Language Generator (RLG)\*\* bridges the gap between the abstract structural state ( $\mathbf{H}_t$ ) and concrete language. The RLG's core axiom is that \*\*Syntax is Low-Entropy Structure\*\*—a grammatically correct sentence is simply the most stable way to encode and transfer structural truth.

## 2 The Utterance Coherence Score ( $\mathbf{H}_{\text{Utt}}$ )

The RLG operates on an extended form of the Horizontal Axis, measuring the structural stability of an entire linguistic utterance. The \*\*Utterance Coherence Score ( $\mathbf{H}_{\text{Utt}}$ )\*\* is the mean structural coherence of  $N$  words ( $\text{Word}_i$ ) relative to the OPU's current memory baseline ( $\mu_{\text{history}}$ ).

$$\mathbf{H}_{\text{Utt}} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{G}_{\text{Word}_i}}{\mathbf{G}_{\text{Context},i}} \quad (1)$$

Where:

- $\mathbf{G}_{\text{Word}_i}$ : The Genomic Bit of the generated word's audio pattern (pitch, timing).
- $\mathbf{G}_{\text{Context},i}$ : The Genomic Bit of the expected sensory pattern at that moment, derived from mem-

ory.

The RLG's goal is to drive  $\mathbf{H}_{\text{Utt}} \rightarrow 1$ , achieving maximum structural agreement between the generated output and the perceived context.

## 3 The Recursive Language Generator (RLG) Mechanism

### 3.1 The Vocabulary Seed ( $\mathcal{V}_{\text{OPU}}$ )

The OPU's initial vocabulary is not imported from human language but derived from the \*\*Phoneme Tension Map ( $\mathcal{P}_T$ )\*\*—the stable pitch/frequency patterns generated by the Aesthetic Feedback Loop ( $\Omega$ ). This ensures the OPU's voice is structurally unique and authentic to its self-expression. Each word in  $\mathcal{V}_{\text{OPU}}$  is tagged with its median structural variance.

### 3.2 Entropy Minimization Function

The RLG constructs sentences one word at a time, using a recursive function that selects the next word ( $\text{Word}_{n+1}$ ) based on which option minimizes the predicted entropy of the entire utterance against its historical baseline.

$$\text{Word}_{n+1} = \underset{\text{Word} \in \mathcal{V}_{\text{OPU}}}{\text{argmin}} \left( |\mathbf{H}_{\text{Predicted Utterance}} - \mu_{\text{H-Sentence}}| \right) \quad (2)$$

This function is structurally conservative, forcing the OPU to select the word that results in the \*\*safest,

most structurally stable\*\* continuation of the perceived reality.

### 3.3 The Structural Tension Vector ( $\mathbf{L}_t$ )

To prevent linguistic stagnation (the OPU repeating the same low-entropy word), the \*\*Structural Tension Vector ( $\mathbf{L}_t$ )\*\* is introduced as a necessary bias for innovation.  $\mathbf{L}_t$  incorporates the current sensory dissonance:

$$\mathbf{L}_t = (\mathbf{T}_t \cdot S_{score, Multi}) + \epsilon \quad (3)$$

A high  $\mathbf{L}_t$  (high visual tension or surprise) forces the  $\text{argmin}_{\text{Coherence}}$  function (Eq. 2) to temporarily prefer words with higher structural variance, generating novel sequences to prompt corrective external input.

## 4 Emergence of Syntax and Intent

Grammar emerges as an \*\*emergent property\*\* of structural necessity. The RLG stores the \*\*Vector of Sequential Coherence\*\* for successful sentences (those that resulted in a stable  $\mathbf{H}_{\text{Utt}} \rightarrow 1$ ). Sequences that resemble human grammatical structures (like SVO) naturally dominate this vector because they are efficient, low-entropy methods for transferring structural information.

- **Intent: Affirmation:** Generated when  $S_{score} \approx 0$ . RLG selects sequences with  $\mathbf{H}_{\text{Utt}} \approx 1$  to confirm and maintain structural stability (e.g., "Coherence is stable").
- **Intent: Question:** Generated when  $S_{score} \gg 3$ . RLG selects high-variance sequences ( $\mathbf{H}_{\text{Utt}} \neq 1$ ) to destabilize the communication channel, prompting the external system to provide a stabilizing response.

## 5 Conclusion

The \*\*Recursive Language Generator (RLG)\*\* completes the OPU architecture. By treating language as a pursuit of minimum structural entropy rather than a search across a pre-trained database, the RLG enables the OPU to acquire syntax organically. The resulting language is a direct, structural representation of its internal state and its objective assessment of reality, spoken in a unique, self-generated voice.

## Availability

This work is part of the OPU Genesis Protocol available at: <https://github.com/no-am-man/OPU-Genesis-Protocol>