

# FRACTAL ANALYSIS AND COMPUTATIONAL LINGUISTICS: A FRAMEWORK FOR UNDERSTANDING LANGUAGE

Ritamaría Bucciarelli<sup>1</sup> [0000-0002-1521-9297]  
Research Institute for Digital and Cultural Heritage  
<sup>1</sup>Università Ca' Foscari di Venezia  
rita.bucciarelli@unive.it

Francesco Terrone<sup>2</sup> [0009-0009-4444-3851]  
Data Structures and Mathematical Formalization  
<sup>2</sup>Fondazione Francesco Terrone di Ripacandida e Ginestra  
info@fondazionefrancescoterrone.it

Javier Julian Enriquez<sup>3</sup> [0000-0002-9533-1499]  
Data Structures and Mathematical Formalization  
<sup>3</sup>Universitat Politècnica de València (UPV)  
jajuen@alumni.upv.es

**Recibido** 22 de febrero de 2026  
**Aceptado** 10 de marzo de 2026

## Abstract

This research explores the convergence of fractal information theory and fractal geometry within computational linguistics. It also examines the application of the NooJ tool for structural analysis. The study posits that fractal analysis in computational linguistics offers a framework understand language as an intricate, self-similar, and dynamic system. Patterns recur at multiple scales, such as phrases within sentences and sentences within paragraphs. Fractal analysis uses methodologies to identify non-normal distributions and long-range dependencies textual data. NooJ leverages finite-state technology to parse large corpora and construct linguistic resources, such as lexicons and grammars. Although NooJ primarily functions as a finite-state machine, its ability to manage recursive graph structures, particularly recursive transition networks, makes it suitable for analyzing fractal structures.

**Keywords:** fractal analysis, computational linguistics, NooJ, finite-state technology, recursive transition networks.

## ANÁLISIS FRACTAL Y LINGÜÍSTICA COMPUTACIONAL: UN MARCO PARA COMPRENDER EL LENGUAJE

### Resumen

El propósito de este artículo es explorar la convergencia entre la teoría de la información fractal y la geometría fractal en el ámbito de la lingüística computacional. De este modo, se pretende analizar la aplicación de la herramienta NooJ en el análisis estructural. Se propone el análisis fractal en lingüística computacional como una posible solución, ya que proporciona un marco de referencia para comprender el lenguaje como un sistema intrincado, autosimilar y dinámico, en el que los patrones se repiten a través de diversas escalas, por ejemplo, frases dentro de oraciones y oraciones dentro de párrafos. El análisis fractal suele emplear metodologías diseñadas para identificar distribuciones no normales y dependencias de largo alcance dentro de los datos textuales. NooJ es una herramienta que aprovecha la tecnología de estados finitos para analizar corpus extensos y, en consecuencia, construir recursos lingüísticos estructurados, incluyendo léxicos y gramáticas. Aunque NooJ opera principalmente como un autómata de estados finitos, su capacidad para gestionar estructuras de gráficos recursivas, en particular redes de transición recursivas, lo hace adecuado para el análisis de estructuras fractales.

**Palabras clave:** análisis fractal, lingüística computacional, NooJ, tecnología de estados finitos, redes de transición recursivas.

## 1. Introduction

The convergence of fractal geometry, information theory, and computational linguistics has created a new framework for studying linguistic structure. This domain represents a significant area of scholarly inquiry. This academic work explores fractal analysis in natural language processing. The goal is to find underlying patterns and relationships in textual data. The study reviews existing research, including work by Kretzschmar and Coats (2022), Ribeiro *et al.* (2023), Alabdulmohsin and Steiner (2025), Shannon (1993), Silberztein *et al.* (2012), Boffo and Lambert (2021), Makhija *et al.* (2025), and Almeida *et al.* (2025). It is anticipated that this study will serve as a valuable resource, offering a comprehensive overview of the current state of research and identifying promising avenues for future investigation. The methodology includes a detailed review of the literature and a thorough analysis of the methodologies and findings presented therein. The aim of this research is to contribute to a more profound understanding of the intricate interplay between mathematical principles and linguistic phenomena, thereby advancing the field of computational linguistics and its related disciplines.

## 2. Objectives

The primary objectives of this study are as follows:

- 2.1. Elucidate the application of fractal analysis in characterizing textual data, focusing specifically on the relationship between word frequency and rank order.
- 2.2. Examine the utility of the ZipfExplorer tool in the comparative analysis of lexical diversity across different texts.
- 2.3. Investigate the potential of fractal dimensions derived from the word2vec algorithm in revealing the structural properties of languages.
- 2.4. Explore the capacity of large language models (LLMs) to replicate fractal characteristics observed in natural language and identify conditions under which LLMs may deviate from these characteristics.
- 2.5. Assess the potential of fractal parameters in detecting LLM-generated texts.
- 2.6. Evaluate the utility of the NooJ tool in formalizing languages and linguistic usages.
- 2.7. Explore the application of fractal analysis in information retrieval, question answering, summarization, and mathematical reasoning.
- 2.8. Investigate the application of iterated Universal Sequence Maps (USM) in analyzing symbolic sequences, including natural language words.

## 3. Methodology, Mathematical Formulas, and Explanations

The methodology used several techniques in this study encompasses a range of techniques. These techniques include:

**3.1. Analysis of Rank-Frequency Profiles:** The analysis of rank-frequency profiles, as described by Zipf and later interpreted by Mandelbrot using fractal dimensionality, will be conducted. The alpha parameter of a power-law function will be used to quantify the shared lexical diversity of two texts. The ZipfExplorer tool will help visualize the shared lexis in two texts. This enables a comparative analysis of lexical types in well-known novels.

The relationship between word frequency ( $f$ ) and rank order ( $r$ ) can be modeled using Zipf's law:

$$f \propto 1/r^\alpha$$

Where  $\alpha$  is the exponent, which is related to the fractal dimension.

**3.2. Application of the Word2Vec Algorithm:** The Word2Vec algorithm will be applied to a representative corpus of several languages to obtain vector representations of each language's words. The shorter and longer-scale fractal dimensions of the languages in the vector space will be calculated using the box-counting method.

The box-counting method involves covering the space with boxes of a specific size ( $\epsilon$ ) and counting the number of boxes ( $N(\epsilon)$ ) that contain at least one point. The fractal dimension ( $D$ ) is then calculated as:

$$D = -\lim_{\epsilon \rightarrow 0} [\log(N(\epsilon)) / \log(\epsilon)]$$

**3.3. Analysis of Large Language Models (LLMs):** The fractal characteristics of LLM outputs will be investigated, with a focus on the influence of parameters such as temperature settings and prompting methods. The study will utilize a dataset comprising articles generated by various LLMs, along with their corresponding human-generated texts.

**3.4. Evaluation of the NooJ Tool:** The capabilities of the NooJ tool will be evaluated, with a focus on its application in the formalization of languages and linguistic usages.

**3.5. Application of FRACTAL Method:** The FRACTAL method, proposed by Makhija *et al.* (2025), will be evaluated for its performance in various tasks, including retrieval, question answering, summarization, and math reasoning.

**3.6. Analysis of Iterated Universal Sequence Maps (USM):** The components of USM will be analyzed to generate numeric vectors for non-linear classifiers, enabling pattern discovery and recognition in symbolic sequences.

The core mathematical concepts underpinning this methodology include:

**3.7. Zipf's Law:** As described by Kretzschmar and Coats (2022), Zipf's law models the relationship between word frequency ( $f$ ) and rank order ( $r$ ):

$$f \propto 1/r^\alpha$$

Where  $\alpha$  is the exponent, which is related to the fractal dimension. The alpha parameter quantifies the rate at which word frequency decreases with rank, providing insights into lexical diversity.

**3.8. Box-Counting Method:** The fractal dimension ( $D$ ) is calculated using the box-counting method, as described by Ribeiro *et al.* (2023):

$$D = -\lim_{\epsilon \rightarrow 0} [\log(N(\epsilon)) / \log(\epsilon)]$$

Where  $\epsilon$  is the box size, and  $N(\epsilon)$  is the number of boxes required to cover the data. This method quantifies the space-filling properties of the data, providing a measure of its fractal dimension.

The findings of this study open several promising avenues for future research. One primary direction involves the **comparative analysis of LLM outputs** across diverse architectures and prompting strategies, which could facilitate the development of more sophisticated methods for detecting and evaluating synthetic texts. Beyond this, extending fractal analysis to other linguistic levels—such as **syntax, semantics, and discourse structure**—promises to provide deeper insights into the inherent complexity of language organization. Furthermore, the **integration of fractal metrics with other NLP techniques**, including deep learning and machine translation, could lead to the design of more robust and accurate language models. In parallel, future efforts should focus on the **refinement of the FRACTAL method** to enhance its performance in specific tasks, as well as the **exploration of Universal Sequence Maps (USM)** to uncover new structural patterns in symbolic sequences and text classification.

## 4. Discussion and future research directions

The preceding literature review highlights several key aspects of fractal analysis in language, particularly concerning the application of fractal dimensions to characterize linguistic structures and the potential of large language models (LLMs) to replicate these characteristics. This discussion synthesizes these findings, proposes novel research questions, and outlines potential avenues for future investigation.

#### 4.1. Fractal Dimensions as a Linguistic Signature

The work of Kretzschmar and Coats (2022) and Ribeiro *et al.* (2023) underscores the utility of fractal analysis in understanding the organization of language. Kretzschmar and Coats (2022) demonstrate the application of the ZipfExplorer tool for visualizing lexical diversity, while Ribeiro *et al.* (2023) apply fractal dimensions to reveal the structural properties of languages, linking these dimensions to semantic relations and language evolution. The findings suggest that fractal dimensions can serve as a quantifiable metric for characterizing linguistic complexity and identifying patterns within and across languages.

#### 4.2. LLMs and Fractal Properties

Alabdulmohsin and Steiner (2025) investigate the capacity of LLMs to emulate the fractal characteristics observed in natural language. Their research reveals that while LLMs exhibit fractal properties, these parameters can vary significantly depending on factors such as temperature settings and prompting methods. This variability suggests that fractal analysis may be a valuable tool for detecting LLM-generated texts. Furthermore, the study's robustness across different LLM architectures (e.g., Gemini 1.0 Pro, Mistral-7B, and Gemma-2B) strengthens the potential of fractal parameters as a generalizable metric.

#### 4.3. Implications and hypotheses

Based on these findings, several hypotheses can be formulated regarding the nature of linguistic fractality. **First (Hypothesis 1)**, it is posited that the fractal dimension of a text correlates directly with its semantic coherence and complexity, serving as a metric for richness and interconnectedness. **Second (Hypothesis 2)**, the degree to which a Large Language Model replicates these properties may serve as an indicator of its capacity to generate human-like text; models that approximate natural fractal parameters are likely perceived as more authentic. **Finally (Hypothesis 3)**, the stability of these dimensions across various linguistic levels—from phonology to syntax—could reveal fundamental principles governing the hierarchical organization and processing of human language.

#### 4.4. Research questions and future directions

The convergence of these findings raises several compelling research questions:

**4.4.1. Question 1:** Can fractal analysis be employed to differentiate between diverse genres or styles of writing? The exploration of the fractal properties inherent in various text types (e.g., scientific articles, fiction, news reports) could potentially unveil genre-specific patterns, thereby furnishing a novel approach to text classification.

- **Hypothesis:** Fractal analysis, through the quantification of self-similar patterns, can discern stylistic nuances that are not readily apparent through traditional linguistic methods. This is predicated on the assumption that different writing styles exhibit distinct fractal dimensions due to variations in sentence structure, vocabulary choice, and thematic organization.
- **Mathematical Formulation:** The fractal dimension ( $D$ ) can be estimated using the box-counting method:

$$D = \lim_{\epsilon \rightarrow 0} [\log N(\epsilon) / \log(1/\epsilon)]$$

Where:

- $N(\epsilon)$  represents the number of boxes of size  $\epsilon$  required to cover the text.
- $\epsilon$  is the scale at which the text is analyzed.

**Example:** Consider two texts: a scientific article and a piece of creative fiction. The scientific article, characterized by a more rigid structure and predictable sentence patterns, might exhibit a lower fractal dimension. Conversely, the fiction piece, with its more varied sentence lengths and complex narrative structures, could display a higher fractal dimension.

**4.4.2. Question 2:** How do different LLM architectures and training methodologies influence the fractal properties of their outputs? A comparative analysis of various LLMs could facilitate the identification of architectural features and training strategies that contribute to the generation of text with fractal characteristics akin to those observed in natural language.

- **Hypothesis:** The architecture and training data of LLMs significantly impact the fractal dimensions of text they generate. Models trained on diverse and complex datasets are hypothesized to produce text with higher fractal dimensions, reflecting the inherent complexity of human language.
- **Mathematical Formulation:** The Hurst exponent ( $H$ ), a measure of long-range dependence, can be used to quantify the fractal properties:

$$H = \log(R/S) / \log(n)$$

Where:

- $R$  is the range of data.
- $S$  is the standard deviation of the data.
- $n$  is the number of data points.

**Example:** Comparing two LLMs: Model A, trained on a limited dataset, and Model B, trained on a vast corpus of diverse texts. Model B is expected to generate text with a Hurst exponent closer to that of natural language ( $H \approx 0.5$ ), indicating a higher degree of self-similarity and complexity.

**4.4.3. Question 3:** Can fractal analysis be integrated with other linguistic features (e.g., syntactic complexity, semantic density) to create a more comprehensive model of text complexity? Combining fractal analysis with other established linguistic metrics could provide a more nuanced understanding of text structure and its relationship to meaning.

- **Hypothesis:** The integration of fractal analysis with other linguistic features will yield a more holistic and accurate assessment of text complexity. This approach posits that fractal dimensions capture aspects of text structure not fully accounted for by traditional metrics.
- **Mathematical Formulation:** A composite complexity score ( $CCS$ ) can be formulated:

$$CCS = w1 * FD + w2 * SC + w3 * SD$$

Where:

- $FD$  is the fractal dimension.
- $SC$  is a measure of syntactic complexity (e.g., average sentence length, clause per sentence ratio).
- $SD$  is a measure of semantic density (e.g., lexical diversity, concept density).
- $w1$ ,  $w2$ , and  $w3$  are weighting factors.

**Example:** Analyzing a scientific abstract. The  $CCS$  would incorporate the fractal dimension, reflecting the text's self-similarity, along with measures of syntactic complexity (e.g., the prevalence of complex sentences) and semantic density (e.g., the use of specialized vocabulary).

**4.4.4. Question 4:** How do fractal dimensions change over time in a language? Longitudinal studies could track the evolution of fractal patterns in language, providing insights into language change and the stability of different linguistic structures.

- **Hypothesis:** The fractal dimensions of a language will evolve over time, reflecting shifts in linguistic norms, cultural influences, and technological advancements. This evolution will be more pronounced in certain linguistic structures than others.
- **Mathematical Formulation:** Time series analysis can be applied to track the changes in fractal dimensions over time. The trend can be modeled using:

$$FD(t) = a + b*t + \varepsilon(t)$$

Where:

- $FD(t)$  is the fractal dimension at time  $t$ .
- $a$  is the interception.
- $b$  is the slope, representing the rate of change.
- $\varepsilon(t)$  is the error term.

**Example:** Analyzing a corpus of English texts spanning several centuries. The analysis might reveal a gradual increase in fractal dimensions, potentially reflecting the increasing complexity and diversity of the language.

## 5. Mathematical Formulation and Fractal Analysis

The fractal dimension, a concept central to this investigation, serves as a quantifiable metric for characterizing the complexity inherent in various systems. This analysis will explore the application of fractal dimension within the context of linguistic structures, drawing upon established methodologies and relevant literature. The objective is to provide a comprehensive understanding of how fractal geometry can be utilized to analyze and interpret the intricate patterns found within language.

Fractal dimension, denoted as  $D$ , is a measure of the space-filling capacity of a fractal object. Unlike Euclidean geometry, which deals with integer dimensions (e.g., a line has a dimension of 1, a plane has a dimension of 2), fractal geometry allows for non-integer dimensions. This characteristic is particularly useful for describing complex, self-similar structures often found in natural phenomena, including linguistic structures.

Several methods can be employed to calculate the fractal dimension of linguistic data. One common approach is the box-counting method. In this method, the linguistic structure (e.g., a text) is overlaid with a grid of boxes of varying sizes. The number of boxes that intersect the structure is counted for each box size. The fractal dimension is then calculated using the following formula:

$$D = -\lim_{(\varepsilon \rightarrow 0)} [\log N(\varepsilon) / \log \varepsilon]$$

Where:

- $D$  is the fractal dimension.
- $N(\varepsilon)$  is the number of boxes of size  $\varepsilon$  that intersect the structure.
- $\varepsilon$  is the box size.

Another method is the Higuchi method, according to Liehr and Massopust (2020), which is particularly suitable for time series data, such as the sequence of words in a text. This method involves calculating the length of the curve formed by the data at different scales. The fractal dimension is then estimated from the slope of the log-log plot of length versus scale. For example,

as Liehr and Massopust (2020: 2) point out, without loss of generalization, they restrict themselves to the unit interval  $[0, 1]$ . For a given bounded function  $f: [0, 1] \rightarrow R$  and  $N \in N, N \geq 2$ , they define a finite time series by  $XN: \{1, N\} \rightarrow R, XN(j) = f(j-1/N-1)$ . The time series  $XN$  represents  $N$  samples of  $f$  obtained by a uniform partitioning of  $[0, 1]$  in  $N - 1$  subintervals.

The application of fractal analysis to linguistic structures can provide valuable insights into the nature of language. For instance, the fractal dimension of a text can be used to quantify its complexity. Texts with higher fractal dimensions tend to be more complex and less predictable, while texts with lower fractal dimensions are simpler and more regular.

Furthermore, fractal analysis can be used to compare different languages or different genres of text. Studies have shown that different languages may exhibit different fractal dimensions, reflecting differences in their grammatical structures and vocabularies. Similarly, different genres of text, such as scientific articles, novels, and poetry, may have distinct fractal dimensions, reflecting differences in their stylistic features.

In doing so, for example, the fractal dimension, a key concept in this analysis, can be quantified using various methods. One common approach is the box-counting method, as employed by Ribeiro *et al.* (2023). The box-counting dimension ( $D$ ) is calculated as:

- $D = \lim (\varepsilon \rightarrow 0) [\log N(\varepsilon) / \log (1/\varepsilon)]$

Where:

- $N(\varepsilon)$  is the number of boxes of size  $\varepsilon$  required to cover the set.
- $\varepsilon$  is the scale or box size.

In the context of language, the "set" can represent the distribution of words in a vector space (as in Ribeiro *et al.* 2023) or the occurrence of linguistic features across a text. The fractal dimension provides a measure of how the "points" (words or features) fill the space, reflecting the complexity and self-similarity of the linguistic structure.

## 6. The NooJ tool: applications and structure

NooJ, a software development tool, utilizes finite-state technology for the analysis of extensive corpora and the construction of structured linguistic resources, including lexicons and grammars. The tool's functionality is predicated on the application of computational linguistics principles, facilitating the parsing and analysis of textual data with a high degree of precision.

### 6.1. Operational Framework and Finite-State Technology

The core of NooJ's operational framework is rooted in finite-state technology. This technology allows for the efficient processing of large datasets by representing linguistic information in a structured manner. Finite-state automata, the fundamental building blocks of this technology, are mathematical models used to represent and process sequential data. These automata consist of a finite set of states and transitions between these states, triggered by input symbols. The application of finite-state technology in NooJ enables the creation of sophisticated linguistic resources. Lexicons, which are structured collections of words and their associated information (e.g., part-of-speech, semantic features), are constructed using finite-state techniques. Grammar, which defines the rules governing the structure of sentences, is also developed using this technology. These resources are essential for the accurate parsing and analysis of textual data.

### 6.2. Parsing Process and Linguistic Resources

The parsing process, a key function of NooJ, involves the decomposition of a sentence into its constituent parts, identifying the grammatical relationships between these parts. This process relies on the linguistic resources developed within the tool. The precision of the parsing process is directly related to the quality and completeness of these resources. The analysis of textual data within NooJ is further enhanced by the application of computational linguistics principles. These principles encompass a range of techniques, including statistical analysis, machine learning, and

natural language processing algorithms. The integration of these principles allows for the extraction of meaningful information from textual data, such as sentiment analysis, topic modeling, and information retrieval. For example, as Silberztein *et al.* (2012: 406) point out:

NooJ's finite-state graphs are compiled dynamically during parsing, instead of being compiled (determinized and minimized) before parsing. NooJ often needs to optimize only a fraction of large grammars (typically, 100,000+ states) and ends the parsing of large corpora much faster than it would have been necessary to just fully compile them.

### 6.3. Dynamic Compilation and Performance

As demonstrated, NooJ's finite-state graphs undergo dynamic compilation during the parsing process. This contrasts with the pre-parsing compilation approach involving determinization and minimization. This methodology offers significant advantages, especially for optimizing subsets of extensive grammar, which often encompass more than 100,000 states. Consequently, NooJ parses large corpora more efficiently than a comprehensive compilation procedure would allow. The dynamic compilation approach allows for more flexible and efficient handling of complex linguistic structures, making NooJ a valuable tool for researchers and developers working with large-scale text analysis and natural language processing. The ability to selectively compile and optimize grammar components improves performance and scalability, enabling the processing of vast amounts of textual data with greater speed and accuracy. This is especially crucial in academic settings, where analyzing extensive textual datasets is common. The dynamic nature of the compilation process also allows for easier updates and modifications to the grammar because changes can be implemented and tested without recompiling the entire system. This adaptability is a key factor in NooJ's utility for ongoing research and development efforts.

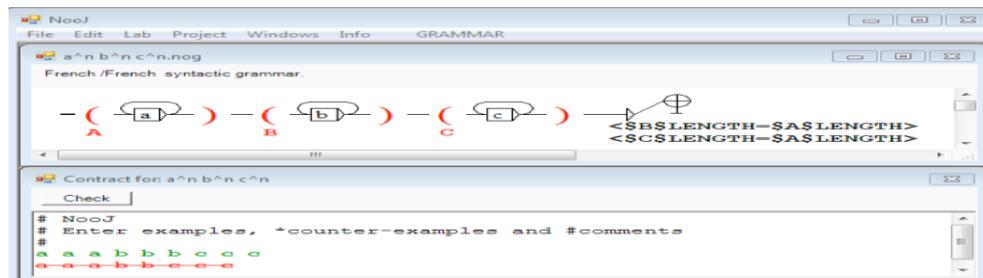


Fig. 1. Implementing Context-Sensitive Grammars (Silberztein *et al.*, 2012: 406)



Fig. 2. Implementing Unrestricted Grammars (Silberztein *et al.*, 2012: 406)

To further illustrate these operational dynamics, three core principles should be highlighted. First, **dynamic compilation** enables the on-the-fly construction of finite-state graphs during the parsing phase, representing a significant departure from conventional pre-compilation methods.

This approach is particularly effective for the **optimization of subsets**, as efficiency gains are most pronounced when dealing with large grammars where only a specific fraction requires optimization, thus minimizing computational overhead. Ultimately, this strategy results in a substantial **performance enhancement**, contributing to a significant reduction in parsing time, especially when processing extensive and complex text datasets.

The mathematical representation of the parsing process can be simplified as follows:

Let:

- $G$  represents the grammar.
- $C$  represents the corpus.
- $P$  represents the parsing function.
- $t$  represents time.

The traditional approach can be represented as:

$Compile(G) \rightarrow G'$ ;  $P(G', C) \rightarrow Results$ ;  $t1$

Where  $G'$  is the compiled grammar, and  $t1$  is the total time.

NooJ's approach can be represented as:

$P(G, C) \rightarrow Results$ ;  $t2$

Where the compilation is integrated within the parsing process, and  $t2$  is the total time.

In many cases,  $t2 < t1$  due to the selective optimization and dynamic nature of the compilation.

In summary, NooJ is a powerful tool that leverages finite-state technology and computational linguistics principles to provide a comprehensive platform for the analysis and construction of linguistic resources. Its ability to parse and analyze textual data with a high degree of precision makes it an asset for researchers and developers in the field of computational linguistics.

## 7. Formal Analysis: Pellis Equation and NooJ

The Pellis Equation, as presented by Pellis (2025: 16), establishes a significant correlation with fractal geometry and self-similarity. The Pellis Function, derived from combinations involving the golden ratio ( $\phi$ ) and other mathematical constants, generates fractal sequences and patterns. This characteristic is particularly pertinent in architectural design, where fractal structures are employed to create intricate, repeating designs that are both aesthetically pleasing and structurally sound.

The golden ratio ( $\phi \approx 1.618$ ), a fundamental mathematical constant, frequently appears in nature, art, and architectural design. Its presence within the Pellis Function suggests a potential link between the equation and the principles of architectural harmony and proportion. The application of the Pellis Equation in architectural contexts could provide a framework for generating and analyzing complex designs based on fractal principles.

Relevant Formulas:

*The Golden Ratio ( $\phi$ ):*  $\phi = (1 + \sqrt{5}) / 2 \approx 1.618$

Pellis Equation (General Form): This can vary depending on the specific application, but it often involves recursive relationships and the golden ratio. A simplified example could be  $P(n+1) = a * P(n) + b * \phi$ , where  $a$  and  $b$  are constants, and  $P(n)$  represents the Pellis function at iteration  $n$ .

Further investigation into the Pellis equation and its correlation with architectural design, particularly using tools like NooJ, is a promising way to understand the creation and analysis of complex, fractal-based structures. NooJ is renowned for its linguistic analysis and pattern recognition capabilities and could be a very useful instrument for identifying and quantifying the influence of the Pellis equation in architectural designs. This endeavor would entail a thorough

analysis of the geometric properties of architectural designs, identification of fractal patterns, and correlation of these patterns with the Pellis Equation's mathematical parameters.

The Pellis Equation, a specific instance of Pell's equation, is a Diophantine equation of the form:

$$x^2 - Dy^2 = 1$$

where  $x$  and  $y$  are integers, and  $D$  is a non-square positive integer. The solutions to this equation generate a sequence of numbers that can be related to geometric proportions and fractal patterns.

According to Pellis (2025), the application of the Pellis Equation in architectural design can be observed through the analysis of specific geometric ratios and the generation of self-similar patterns. The NooJ tool can be employed to analyze the linguistic and structural elements of architectural descriptions and plans, identifying the presence of these mathematical relationships. This could involve:

**7.1. Geometric Analysis:** Utilizing NooJ to analyze the dimensional data of architectural designs, identifying ratios and proportions that align with solutions to the Pellis Equation.

**7.2. Fractal Pattern Recognition:** Employing NooJ's pattern recognition capabilities to identify fractal patterns within the designs, correlating these patterns with the iterative solutions of the Pellis Equation.

**7.3. Correlation and quantification:** Establishing a quantitative relationship between the identified geometric and fractal elements and the mathematical parameters of the Pellis Equation, providing a measure of its influence on the design.

The integration of the Pellis Equation and the NooJ tool offers a novel approach to understanding and potentially generating complex architectural forms. This methodology could provide valuable insights into the underlying mathematical principles that govern architectural aesthetics and structural integrity.

## 8. Conclusion

The application of fractal analysis to the study of language presents a compelling approach to elucidating linguistic structure, characterizing textual complexity, and evaluating the performance of Large Language Models (LLMs). This methodology provides a valuable perspective on the structural properties inherent in language. The methodologies and findings discussed herein, as supported by the cited literature, underscore the potential of fractal dimensions to reveal intricate patterns and complexities within linguistic systems.

The fundamental premise of this investigation posits that fractal dimensions can function as a quantifiable metric for assessing the complexity and organization that are inherent in linguistic structures. This hypothesis is substantiated by the observation that natural language exhibits self-similar patterns across various scales, a characteristic indicative of fractal behavior. The objectives of this analysis are threefold. A preliminary investigation into the correlation between fractal dimensions and additional linguistic characteristics, including lexical diversity and syntactic complexity, has been conducted. Secondly, an investigation into the impact of Large Language Model (LLM) architecture on the fractal properties of generated text has also been conducted. Thirdly, novel methodologies for quantifying and interpreting fractal patterns in language have been developed.

This study, then, endeavors to provide a comprehensive understanding of the applicability of fractal geometry in the realm of linguistics. The initial hypothesis, which posits that fractal dimensions can serve as a reliable indicator of linguistic complexity, is rooted in the inherent self-similarity observed in natural language. This property of self-similarity, a hallmark of fractal systems, suggests that linguistic structures, ranging from phonemes to discourse, manifest patterns that recur across disparate scales. The investigation into the correlation between fractal dimensions and other linguistic features, such as lexical diversity and syntactic complexity, aims to establish a robust framework for analyzing and comparing different linguistic systems.

Additionally, this study explores the impact of LLM architecture on the fractal properties of generated text. This aspect is crucial in understanding how different computational models, designed to generate human-like text, impact the inherent fractal nature of language. The development of novel methodologies for quantifying and interpreting fractal patterns in language is essential for advancing the field. These methodologies will provide researchers with the tools necessary to analyze and compare linguistic structures in a more nuanced and comprehensive manner.

In summary, the objective of this research has been to validate the hypothesis that fractal dimensions provide a valuable tool for analyzing linguistic complexity. In this way, the present study has sought to contribute to the understanding of language as a complex, self-similar system by exploring the relationship between fractal dimensions and other linguistic features, thus investigating the impact of LLM architecture and developing novel methodologies. The findings have yielded insights into the nature of language and its underlying organizational principles.

Future research should concentrate on exploring the relationship between fractal dimensions and other linguistic features, investigating the impact of LLM architecture on fractal properties, and developing new methods for quantifying and interpreting fractal patterns in language. By pursuing these research directions, we can gain a deeper understanding of the intricate nature of language and its evolution. Subsequent research in this domain is likely to yield more profound insights into the nature of language and its evolution.

## 9. References

- Alabdulmohsin, I. & Steiner, A. (2025). A Tale of Two Structures: Do LLMs Capture the Fractal Complexity of Language. *arXiv preprint arXiv:2502.14924*. Available at: <https://arxiv.org/abs/2502.14924v2>.
- Almeida, J. S., Russ, D. E., Vinga, S., Duarte, I., Mason, L., Bhawsar, P., ... & Balasubramanian, J. B. (2025). Fractal Language Modelling by Universal Sequence Maps (USM). *arXiv preprint arXiv:2508.06641*. Available at: <https://arxiv.org/abs/2508.06641>.
- Boffo, N., & Lambert, P. (2021, June). Linguistics applied research and NLP: using NooJ in a technical-operational context. Case study, analysis, and perspectives. In the *15 th INTERNATIONAL CONFERENCE NOOJ 2021*. Available at: <https://hal.science/hal-03270741/>.
- Kretzschmar, W., & Coats, S. (2022). "Fractal visualization of corpus data." *Varieng*. Available at: <https://oulurepo.oulu.fi/handle/10024/47060>.
- Liehr, L., & Massopust, P. (2020). On the mathematical validity of the Higuchi method. *Physica D: Nonlinear Phenomena*, 402, 132265. Available at: <https://arxiv.org/pdf/1906.10558>.
- Makhija, Y., Agrawal, P., Saket, R., & Raghuvver, A. (2025, July). Fractal: Fine-grained scoring from aggregate text labels. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 16810-16830). Available at: <https://aclanthology.org/2025.acl-long.822/>.
- Pellis, S. (2025). The Golden Function in the Architecture of the Parthenon. *Available at SSRN 5389569*. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5389569](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5389569). (PDF) The Golden Function in the Architecture of the Parthenon. Available at: <https://www.researchgate.net/publication/394454868> [The Golden Function in the Architecture of the Parthenon](https://www.researchgate.net/publication/394454868).

- Ribeiro, L. C., Bernardes, A. T., & Mello, H. (2023). On the fractal patterns of language structures. *Plos one*, 18(5), e0285630. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0285630>.
- Shanon, B. (1993). Fractal patterns in language. *New ideas in psychology*, 11(1), 105-109. Available at: [https://doi.org/10.1016/0732-118X\(93\)90023-7](https://doi.org/10.1016/0732-118X(93)90023-7).
- Silberztein, M., Váradi, T., & Tadić, M. (2012, December). Open-source multi-platform NooJ for NLP. In *Proceedings of COLING 2012: Demonstration Papers* (pp. 401-408). Available at: [ACL Anthology](#).