

# Comparative Symbolic Grammar Analysis of Fungal Electrical Networks: Methods and Data

---

**Zubair Chowdhury**

Independent Researcher

Wimbledon, England

January 2, 2026

---

## Abstract

We present a comparative preprocessing and symbolic encoding methodology for the analysis of fungal electrical recordings across multiple species. Electrophysiological data from four species recorded by Adamatzky (2021) were integrated with a 137-hour high-resolution *Schizophyllum commune* recording from the author's laboratory. All datasets were z-score normalised, resampled to a common length of 50,000 samples, and quantised into a five-symbol alphabet (A–E) to enable fair cross-species comparison. Shannon entropy of the resulting symbol distributions revealed striking differences in syntactic richness: three Adamatzky species showed near-degenerate entropies (0.0003–0.016 bits), *Flammulina velutipes* showed moderate diversity (0.50 bits), and the author's fast-spike *S. commune* dataset reached maximum information density (2.32 bits). This methods paper documents the pipeline and motivates subsequent grammar-based compression and hierarchical structure analysis.

---

## 1. Data Sources

Electrophysiological recordings from four fungal species were obtained from the open Zenodo repository "Recordings of electrical activity of four species of fungi" (record 5790768, Adamatzky 2021):

- *Cordyceps militaris* (1,900,145 samples; 108.76 MB)
- *Flammulina velutipes* (Enoki mushroom; 1,210,938 samples; 57.35 MB)
- *Omphalotus nidiformis* (Ghost fungi; 3,279,569 samples; 156.42 MB)
- *Schizophyllum commune* (4-species dataset; 263,959 samples; 13.54 MB)

In addition, a fast-spike window of 3,534 samples extracted from a 137-hour continuous *Schizophyllum commune* recording acquired in the author's laboratory (Chowdhury 2025 dataset) was included as an independent high-resolution control (0.13 MB).

All differential electrode channels were exported as tab-delimited text files and imported into Python 3.14 (pandas 2.3, NumPy 2.4) for processing.

---

## 2. Signal Preprocessing and Normalisation

For each dataset, only numeric columns (voltage measurements in millivolts) were retained, and the first numeric channel was selected as the reference signal for comparative analysis. This ensured consistency across species with different numbers of electrode pairs.

Each selected channel was z-score normalised according to:

$$\$z(t) = \frac{V(t) - \mu}{\sigma}$$

where  $V(t)$  is the instantaneous voltage,  $\mu$  is the channel mean, and  $\sigma$  is the standard deviation (computed with Bessel's correction,  $N-1$  denominator).

This normalisation centres and scales each dataset to zero mean and unit variance, removing any offset or amplitude bias introduced by different electrode geometries or amplifier gains.

---

### 3. Temporal Resampling to Common Length

Because the Adamatzky datasets span periods of 2–3 million samples (varying recording durations and sampling rates), and the Chowdhury fast-spike window contains only 3,534 samples, direct comparison of raw time series is not feasible. To enable fair cross-species analysis, each normalised trace was resampled to a common length of 50,000 samples by linear interpolation using NumPy's `interp` function:

$$\$V'_{\text{resampled}}(i) = \text{interp}\left(x[i], x[\text{old}], V_z\right), \quad i = 1, \dots, 50000$$

where  $x[\text{old}]$  represents the original time indices (normalised to [0, 1]),  $x'$  represents the new equispaced grid, and  $V_z$  is the z-scored signal.

This approach preserves the shape and gross dynamics of each time series while bringing all datasets into a common temporal reference frame for downstream comparison.

---

### 4. Symbolic Quantisation into Five-Symbol Alphabet

The resampled, normalised traces were then discretised into a five-symbol alphabet (A, B, C, D, E) using empirical quintiles of the amplitude distribution. For each dataset, the z-scored values were divided into five equal-probability bins defined by the 20th, 40th, 60th, and 80th percentiles:

- Bin 1 ( $z < Q_{20}$ ): **A**
- Bin 2 ( $Q_{20} \leq z < Q_{40}$ ): **B**
- Bin 3 ( $Q_{40} \leq z < Q_{60}$ ): **C**
- Bin 4 ( $Q_{60} \leq z < Q_{80}$ ): **D**
- Bin 5 ( $z \geq Q_{80}$ ): **E**

Each continuous sample was assigned the corresponding symbol, yielding one long discrete string of length 50,000 per dataset.

This choice of five symbols reflects a balance between coarse-graining (loss of detail) and alphabet size (combinatorial explosion). The quintile-based quantisation ensures that each symbol appears with equal expected frequency in a uniform distribution, allowing differences in symbol entropy to reflect genuine structure in the data rather than arbitrary bin widths.

---

### 5. Syntactic Richness Quantification: Shannon Entropy

As a first-pass proxy for the syntactic complexity and information content of each symbolic sequence, the Shannon entropy of the symbol distribution was computed:

$$\text{H} = -\sum_{s \in \{A,B,C,D,E\}} p(s) \log_2 p(s)$$

where  $p(s)$  is the relative frequency of symbol  $s$  in the sequence.

The theoretical minimum is  $H = 0$  (all samples map to a single symbol), and the theoretical maximum for a five-symbol alphabet is  $H = \log_2(5) \approx 2.322$  bits, achieved when all symbols appear with equal probability (maximum entropy / minimum redundancy).

---

## 6. Results: Entropy Across Species

The Shannon entropy of the five-symbol sequences is summarised in Table 1:

Dataset	Original Length	Resampled Length	Symbols Used	Entropy (bits)	% of Max
<i>Cordyceps militaris</i> (Adamatzky 2021)	1,900,145	50,000	1–2	0.00064	0.03%
<i>Flammulina velutipes</i> (Adamatzky 2021)	1,210,938	50,000	4–5	0.49883	21.5%
<i>Omphalotus nidiformis</i> (Adamatzky 2021)	3,279,569	50,000	1–2	0.00034	0.01%
<i>Schizophyllum commune</i> (Adamatzky 2021)	263,959	50,000	2–3	0.01586	0.68%
<i>Schizophyllum commune</i> (Chowdhury 2025, fast-spike)	3,534	50,000	4–5	2.32193	99.9%

Three of the four Adamatzky species—*C. militaris*, *O. nidiformis*, and the Adamatzky *S. commune*—exhibit extremely low entropy, indicating that the resampled quantised sequences are dominated by one or two symbols and are thus syntactically degenerate at the five-symbol level. *Flammulina velutipes* shows moderate diversity (21.5% of maximum), with four or five symbols present. By contrast, the Chowdhury fast-spike *S. commune* dataset achieves near-maximal entropy (99.9%), with all five symbols appearing with nearly equal frequency—a stark difference that suggests the fast-spike window captures genuinely random or maximally diverse electrical activity compared to the lower-frequency recordings in the Adamatzky dataset.

---

## 7. Interpretation and Next Steps

The large disparity in symbol entropy across species and recording modalities motivates further investigation using hierarchical and context-sensitive grammar models. The near-zero entropy of three Adamatzky datasets may reflect either genuine biological constraints (e.g., sparse, repetitive electrical patterns in mature mycelial networks) or an artefact of the resampling-and-quantisation pipeline applied to low-information-density signals. Conversely, the maximal entropy of the fast-spike *S. commune* recording may indicate either high biological signal richness or the capture of noise and high-frequency transients that increase apparent randomness.

Future analyses will apply context-free grammar induction (e.g. Sequitur-like compression algorithms) and Lempel-Ziv-Welch (LZW) compression ratios to quantify hierarchical rule structure and redundancy in each sequence. These methods are expected to reveal whether the observed entropy differences translate into differences in compressibility and thus in true syntactic complexity—a key question for determining whether fungal electrical networks exhibit learnable, grammar-like information processing.

---

## References

Adamatzky, A. (2021). Recordings of electrical activity of four species of fungi. *Zenodo*.  
<https://doi.org/10.5281/zenodo.5790768>

Chowdhury, Z. (2025). 137-hour continuous *Schizophyllum commune* fast-spike recording. Unpublished laboratory data.

NumPy Developers. (2024). NumPy 2.4. <https://numpy.org/>

pandas Development Team. (2024). pandas 2.3. <https://pandas.pydata.org/>

---

**Date submitted:** January 2, 2026

**Corresponding author:** Zubair Chowdhury ([zubair.chowdhury.uk@gmail.com](mailto:zubair.chowdhury.uk@gmail.com))  
(<https://grumpyjournalist.substack.com/>)