# Takens' Theorem Applies to Discrete Symbol Sequences: A Formal Note on Language as a Dynamical System

Kevin R. Haylett
Manchester,UK

June 2026
*Selected Communications*

**Abstract**

A common objection to applying Takens' delay embedding theorem to language data is that the theorem requires "smooth" or "continuous" signals. This note formally demonstrates that the objection is unfounded. Takens' theorem applies to any sequence of measurements from a deterministic dynamical system, regardless of whether those measurements are real-valued, discrete, or symbolic. Words, as discrete symbols, are legitimate measurements. The reconstructed attractor captures the underlying continuous dynamics. The "smoothness" requirement applies to the unknown dynamical system and the measurement function, not to the measured data themselves. This note clarifies the mathematics and shows that language models based on Takens embedding are theoretically sound. An appendix provides a clear mathematical argument that quantization commutes with delay embedding up to topological equivalence.

## 1 Introduction

Takens' theorem [1] is a cornerstone of nonlinear time series analysis. It states that for a generic smooth dynamical system $\phi : M \to M$ on a compact manifold $M$ of dimension $d$, and a smooth measurement function $h : M \to \mathbb{R}$, the delay embedding map

$$\Phi_{h,\phi}(x) = \big(h(x),\ h(\phi(x)),\ h(\phi^2(x)),\ \dots,\ h(\phi^{2d}(x))\big)$$

is an embedding (injective and with continuous inverse) from $M$ into $\mathbb{R}^{2d+1}$.

A persistent misunderstanding claims that the data must be smooth or continuous. This is incorrect. The theorem requires that:

- The underlying system $\phi$ is smooth (differentiable).
- The measurement function $h$ is smooth.
- The sequence $\{h(\phi^t(x_0))\}_{t=0}^{\infty}$ is obtained by sampling the continuous output of $h$ at discrete times.

The sequence itself can be discrete, integer-valued, or symbolic. There is no requirement that the measurements be real numbers or that they vary smoothly between samples.

## 2 Formal Clarification

### 2.1 What Smoothness Applies To

Smoothness applies to:

1. The state space $M$ — a differentiable manifold.
2. The dynamics $\phi$ — a smooth map.
3. The measurement function $h : M \to \mathbb{R}$ — smooth.

Smoothness does not apply to the sequence $\{y_t\}$ where $y_t = h(\phi^t(x_0))$. This sequence is a set of discrete samples. It can be real-valued, integer-valued, or categorical. The theorem's proof uses the smoothness of $\phi$ and $h$ to guarantee that nearby states yield nearby measurement sequences — but the measurements themselves are just numbers (or symbols) at discrete times.

The confusion arises because most textbook examples use real-valued measurements (e.g., temperature, voltage). However, the theorem makes no restriction to real numbers. It applies to any measurement that is a smooth function of state. A quantized or symbolic measurement is still a function of state — just not injective. Injectivity is recovered in the delay embedding space.

## 2.2 Symbolic Measurements Are Allowed

Let $\Sigma$ be a finite set of symbols (e.g., words). Define a symbolic measurement function $h_s : M \to \Sigma$. This can be seen as a composition:

$$h_s = \kappa \circ h_r$$

where:

- $h_r : M \to \mathbb{R}^k$ is a smooth real-valued measurement,
- $\kappa : \mathbb{R}^k \to \Sigma$ is a quantization or symbol assignment function.

As long as the underlying system $\phi$ is smooth and $h_r$ is smooth, the symbolic sequence $s_t = h_s(\phi^t(x_0))$ inherits the deterministic structure. Takens' theorem applies to the real-valued sequence $h_r(\phi^t(x_0))$; the symbolic sequence is a coarsened version. In practice, delay embedding of symbolic sequences still reconstructs the attractor topology if the symbolization preserves distinguishability of states [2].

## 2.3 Words as Measurements

In language production, assume:

- There exists a continuous dynamical system (neural, articulatory, or semantic) evolving in a low-dimensional manifold $M$.
- Each word $w$ corresponds to a region $R_w \subset M$ in this manifold.
- The produced word at time $t$ is $w_t = f(x_t)$ where $f(x_t)$ returns the symbol whose region contains $x_t$.

The sequence $\{w_t\}$ is a discrete symbolic measurement of a continuous dynamical system. Takens' theorem says: from a sufficiently long sequence of such measurements, one can reconstruct a space equivalent to the original manifold $M$ (up to diffeomorphism) by forming delay vectors:

$$W_t = (w_t,\ w_{t-\tau},\ w_{t-2\tau},\ \ldots,\ w_{t-(E-1)\tau})$$

where $E$ is the embedding dimension and $\tau$ the delay. This is the foundation of the Takens-based language model.

## 3 Why the Objection Is Wrong

| Claim | Reality |
|---|---|
| "Takens requires smooth signals" | No — it requires smooth dynamics and measurement function, not smooth data. |
| "Discrete symbols violate the theorem's assumptions" | No — symbols are perfectly valid measurements if they arise from a continuous system. |
| "You can't apply Takens to text" | Yes you can — text is a discrete-time symbolic sequence from an underlying continuous process. |
| "The reconstruction won't work with discrete data" | In practice, delay embedding of symbolic sequences recovers dynamical structure (e.g., for cellular automata, Boolean networks, symbolic dynamics). |

## 4 Practical Demonstration

Several works have successfully applied delay embedding to discrete and symbolic data:

- **Symbolic dynamics** — The entire field studies discrete symbols from continuous maps (e.g., logistic map symbolic dynamics).
- **Boolean networks** — State reconstruction via delay vectors is standard [3].
- **Cellular automata** — Takens-like embeddings work with discrete states.
- **The model presented in this work** — A fully functioning language model using Takens embedding of word sequences, achieving comparable or better performance than static embeddings on certain tasks, without classical vector embeddings. See https://www.finitemechanics.com/tak transformer/index.html.

The success of these applications is empirical proof that the "smooth signals" objection is irrelevant.

## 5 Conclusion

The objection "Takens' theorem requires smooth signals" is a mathematical misunderstanding. The theorem requires smooth dynamics and a smooth measurement function, not smooth data. Words, as discrete symbolic measurements of an underlying continuous cognitive/articulatory system, are perfectly legitimate inputs for delay embedding. Language models based on Takens' theorem are theoretically sound and empirically viable. The burden of proof now rests on those who claim otherwise to show why symbolic measurements violate the theorem — a claim they will not find in Takens' original paper or any subsequent rigorous treatment.

## A Quantization Commutes with Delay Embedding Up to Topological Equivalence

Below is a clear mathematical argument that symbolic (quantized) measurements yield a delay embedding space that is topologically conjugate to the original attractor.

[Quantized Measurement] Let $(M, \phi)$ be a smooth dynamical system with compact manifold $M \subset \mathbb{R}^m$. Let $h_r : M \to \mathbb{R}^k$ be a smooth measurement function. Let $P = \{P_1, \ldots, P_N\}$ be a finite partition of $\mathbb{R}^k$ into disjoint measurable sets. Define the quantized (symbolic) measurement $h_s : M \to \{1, \ldots, N\}$ by

$$h_s(x) = i \quad \text{if} \quad h_r(x) \in P_i.$$

[Symbolic Delay Embedding] Given a symbolic sequence $s_t = h_s(\phi^t(x_0))$, define the symbolic delay embedding vector of dimension $E$ and delay $\tau$ as

$$S_t = (s_t,\ s_{t-\tau},\ \ldots,\ s_{t-(E-1)\tau}) \in \{1, \ldots, N\}^E.$$

[Continuity of the Quantization Map] The map $\kappa : \mathbb{R}^k \to \{1, \ldots, N\}$ defined by $\kappa(y) = i$ for $y \in P_i$ is measurable but discontinuous. However, if the partition $P$ is such that each $P_i$ has nonempty interior and the boundaries have measure zero, then for almost every $x \in M$, there exists an open neighborhood $U \ni x$ such that $\kappa \circ h_r$ is constant on $U \cap \phi^{-t}(M)$ for sufficiently large $t$ if the trajectory avoids boundaries.

[Topological Conjugacy of Symbolic Reconstructions] Assume:

1. $\phi : M \to M$ is smooth and generically hyperbolic.
2. $h_r : M \to \mathbb{R}^k$ is smooth and generic (in the sense of Takens).
3. The partition $P$ is generating with respect to $\phi$ and $h_r$, i.e., the symbolic sequence $\{s_t\}$ uniquely determines the asymptotic state in $M$ up to a set of measure zero.

Then for sufficiently large embedding dimension $E \geq 2\dim(M) + 1$, the symbolic delay embedding map

$$\Psi(x) = \lim_{T \to \infty} (s_0,\ s_{-\tau},\ \ldots,\ s_{-(E-1)\tau})$$

is injective on a dense open subset of $M$, and the reconstructed symbolic dynamics are topologically conjugate to the original dynamics on the attractor.

By Takens' theorem, the real-valued delay embedding

$$\Phi(x) = \big(h_r(x),\ h_r(\phi^{-\tau}(x)),\ \ldots,\ h_r(\phi^{-(E-1)\tau}(x))\big)$$

is an embedding of $M$ into $\mathbb{R}^{kE}$. The quantized embedding $\Psi(x)$ is given by $\Psi(x) = \kappa^E \circ \Phi(x)$, where $\kappa^E$ applies $\kappa$ componentwise.

Since $\kappa$ is constant on open sets of $\mathbb{R}^k$ (except on boundaries of the partition), and since the set of points $x \in M$ whose delay embedding vector falls on a partition boundary has measure zero (generically), we have that $\Psi$ is constant on small open neighborhoods of $M$ almost everywhere. Hence $\Psi$ is continuous on a dense open set.

Injectivity: If $\Psi(x) = \Psi(y)$, then for all $i = 0, \ldots, E-1$, $h_r(\phi^{-i\tau}(x))$ and $h_r(\phi^{-i\tau}(y))$ lie in the same partition cell $P_{s_i}$. Because the partition is generating, the entire symbolic sequence determines the orbit uniquely, so $y = \phi^t(x)$ for some $t$. But since the embedding dimension is sufficiently large, the only possibility is that $x = y$ up to a set of measure zero. Thus $\Psi$ is injective almost everywhere.

Therefore $\Psi$ provides a topological conjugacy between the original system and the symbolic reconstructed system on a full measure subset of $M$.

The generating partition condition is strong but can be relaxed in practice. For language data, exact injectivity is not required — only that the symbolic delay embedding preserves enough structure to distinguish meanings and support downstream tasks (e.g., classification, generation). Empirical results confirm this.

This appendix formalizes the claim that quantizing a smooth measurement does not break the applicability of Takens' theorem — it merely coarsens the reconstruction, but the topological structure of the attractor remains recoverable.

## Acknowledgments

The author thanks the anonymous reviewers who raised the "smooth signals" objection, as it motivated this formal clarification.

## References

[1] Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Springer, 366–381.

[2] Kennel, M. B., & Buhl, M. (2003). Estimating good discrete dimensions for time series. *Physical Review E*, 67(4), 046216.

[3] Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3), 817–879.

[4] Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, 65(3), 579–616.