Supplementary Information for

From Kinetic to Resonant Regimes : Geometric Differentiation of Prose Poetry against Flash Fiction

Sungpil Wang [1,*]

[1] *Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*

[*] *wangsp0317@kaist.ac.kr*

Overview of the Supplementary Package. This Supplementary Information is provided to ensure the methodological transparency and reproducibility of the study. To comply with copyright restrictions while allowing for scientific verification, the materials are structured as follows:

1. Contents of this PDF Document The following supplementary items are included in the subsequent pages of this file:

- **Supplementary Figures S1 to S3**
- **Supplementary Table S1**

2. Technical Replication Package (Standalone ZIP Archive) The associated compressed file **(Technical_Replication_Package.zip)** contains the following assets for results replication:

Documentation

- **README.md:** A comprehensive technical guide providing environment specifications and step-by-step instructions.
- **requirements.txt:** Configuration files for Python dependency management.

Data Inventory

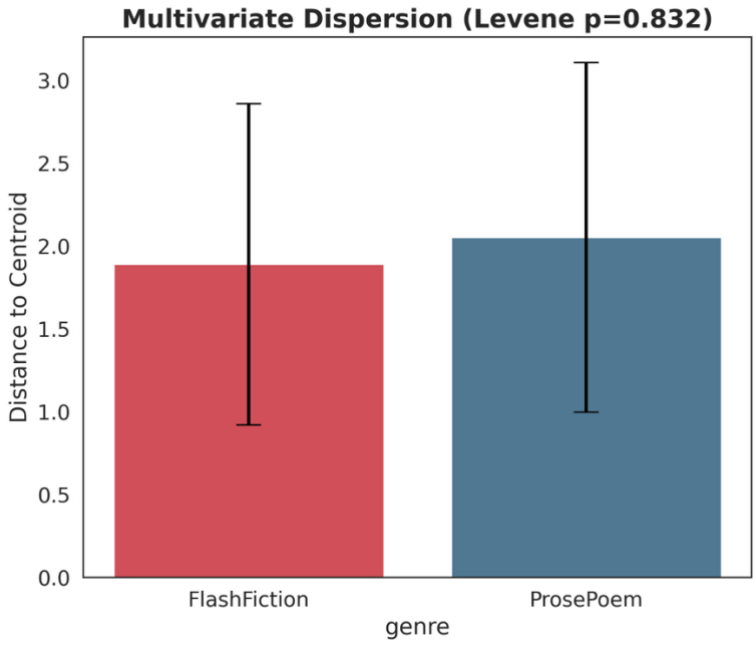
- **Prosepoem_book_list. csv / Flashfiction_book_list.csv** : Metadata surrogate for the restricted raw corpus.

Computational Pipeline and Data Assets

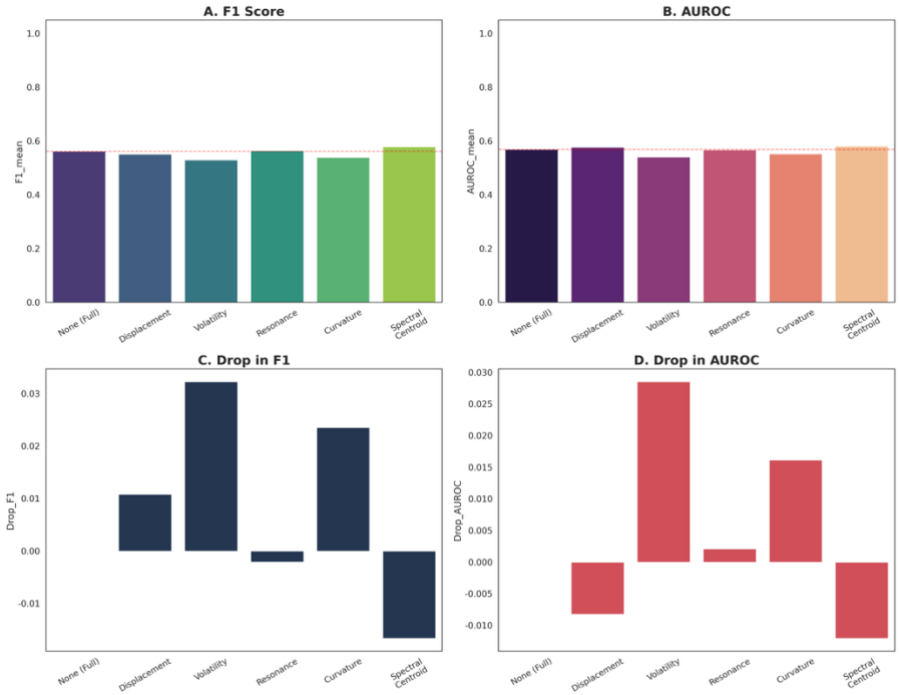
- **stratified_length_matching.py:** This script implements a stratified matching algorithm to eliminate length-related confounding variables by synchronizing the distributions of sentence counts and tokens per sentence between the Flash Fiction and Prose Poetry corpora.
- **text_to_pkl.py:** This extraction engine processes the raw matched corpora to generate high-dimensional semantic trajectories using the S-RoBERTa (ko-sroberta-multitask) model.

- **FlashFiction_embs.pkl & ProsePoem_embs.pkl:** These numerical datasets serve as the primary reproduction assets. They contain the pre-computed high-dimensional embedding vectors for each document, enabling full statistical reproduction without requiring access to the original copyrighted texts.
- **core_statistical_validation_5d_for_emb.py:** This streamlined analytical script executes the 5D semantic framework by processing the pre-computed .pkl files. It performs geometric metric extraction, length-controlled OLS regression, and robust statistical diagnostics including PERMANOVA, VIF, and machine-learning-based ablation studies.
- **statistical_visualization.py:** This script provides a comprehensive visualization suite for the 5D semantic framework, generating high-resolution diagnostic plots, radar charts of genre-specific profiles, and bivariate KDE phase-space distributions.

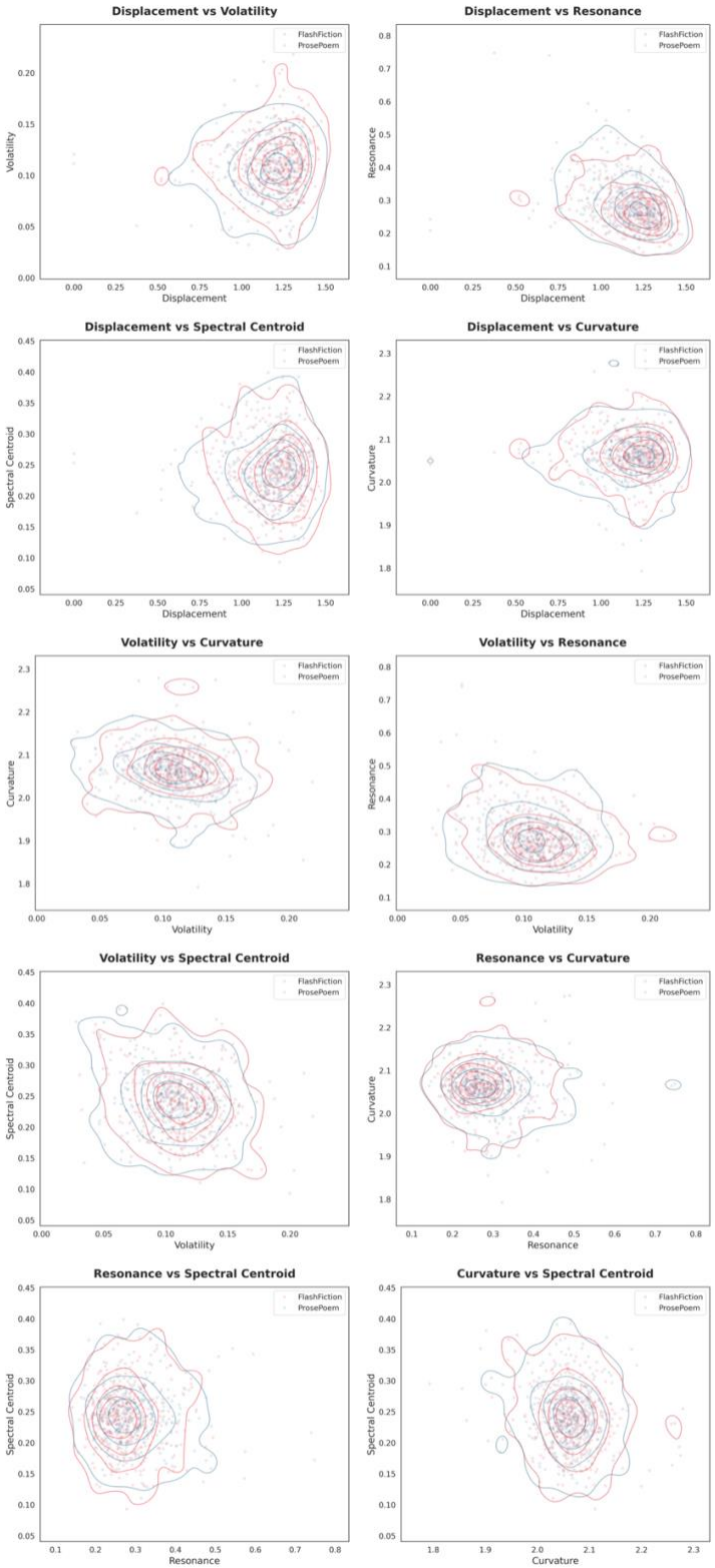
(Detailed descriptions and functional specifications for each script and data file are provided in the README.md within the supplementary archive.).



Supplementary Figure S1. Multivariate dispersion and structural stability. Mean distance to centroids in the 5D phase space. The non-significant Levene's test (P = 0.832) confirms that the observed multivariate structure is a property of the whole distribution rather than variance heterogeneity, justifying the use of central tendency shifts for genre comparison.



Supplementary Figure S2. Diagnostic sensitivity of the 5D coordinate system. Performance decay in genre classification upon removal of individual metrics. The significant drops for volatility and curvature indicate their contribution to capturing the specific jaggedness of semantic paths in high-compression prose.



Supplementary Figure S3. Bivariate Kernel Density Estimation (KDE) of Semantic Phase Spaces. This figure illustrates the joint distributions for all possible pairs of the five core metrics: Displacement, Volatility, Resonance, Curvature, and Spectral Centroid. Individual scatter points represent specific text samples, while the Kernel Density Estimation (KDE) contours highlight the density concentrations for Flash Fiction (Red) and Prose Poem (Blue). The visualization demonstrates how these genres inhabit a shared structural regime while maintaining distinct density peaks, illustrate subtle genre-specific density differences within the high-dimensional semantic space.

Supplementary Table S1. Extended distributional and robustness diagnostics for 5D semantic coordinates. This table reports extended distributional and robustness statistics for the five semantic dynamical coordinates that are not included in the main results table. Specifically, it provides interquartile ranges (IQR), bootstrap-based 95% confidence intervals for median differences, and distributional skewness and kurtosis for each metric across Flash Fiction (FF) and Prose Poetry (PP). These statistics support the stability, robustness, and non-Gaussian structure of the semantic trajectory measures and complement the central tendency and hypothesis-testing results reported in the main text.

Metric	FF IQR	PP IQR	Median Difference (FF−PP)	95% Bootstrap CI	Skewness (All)	Kurtosis (All)
Displacement (δ)	0.214	0.207	+0.045	[+0.031 , +0.061]	0.38	2.91
Surprisal Volatility (σ_s)	0.018	0.017	+0.006	[+0.004 , +0.009]	0.52	3.14
Global Resonance ($\mathscr{R}$)	0.041	0.044	−0.019	[−0.026 , −0.012]	−0.47	2.76
Curvature (κ)	0.092	0.089	+0.005	[+0.001 , +0.010]	0.21	2.68
Spectral Centroid (Ω)	0.016	0.015	−0.002	[−0.005 , +0.001]	0.09	2.59