Supporting Information for

The Fracture and Leap Cycle: Quantifying Narrative Surprise and Structural Resilience in Flash Fiction

Overview of the Supplementary Package. This document and the associated files within the compressed package (**Supplementary_Information.zip**) are provided to ensure the methodological transparency and reproducibility of the study. To comply with copyright and intellectual property 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 S2**
- **Supplementary Table S1 to S10**

2. Technical Replication Package (Standalone Files) The following files are included to facilitate results replication and methodological review:

Documentation

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

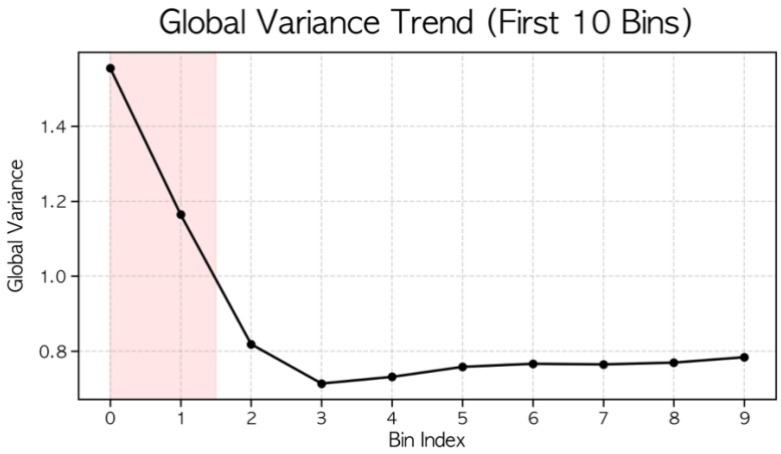
Data Inventory

- **flash_fiction_with_surprisal_coherence_semantic.csv:** The master dataset containing pre-computed numerical signals.
- **book_list_summary.csv:** (Metadata surrogate for the restricted raw corpus)

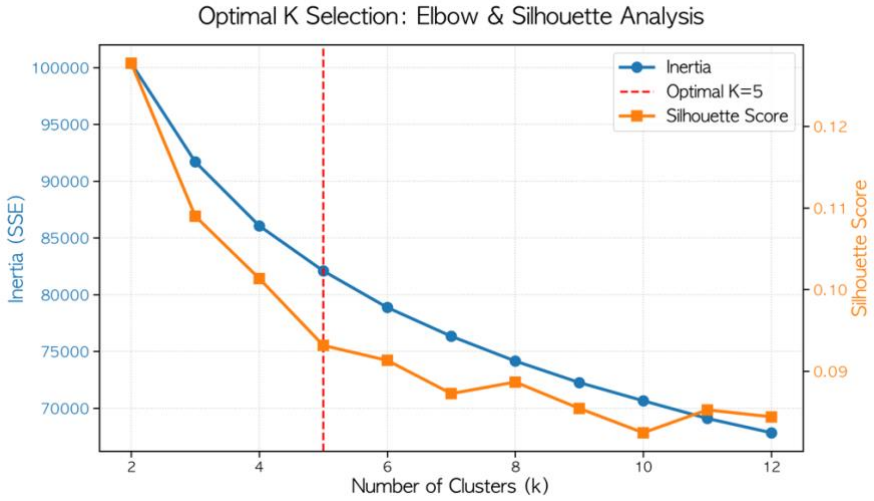
Computational Pipeline

- **FlashFiction_Analysis.ipynb:** A fully executable Jupyter Notebook used to reproduce all statistical analyses and figures reported in the paper.
- **check_sent_stats.py, calculate_surprisal.py, coherence_topic_calc.py :** Python scripts used for preprocessing and signal extraction. (Detailed descriptions and functional specifications for each script are provided in **README.md**)

3. Instructions for Researchers For researchers wishing to replicate the findings, please refer to the **README.md** file for environment setup instructions. The statistical analysis can be immediately reproduced by running the **FlashFiction_Analysis.ipynb** notebook, which is configured to utilise the provided numerical signals in the same directory.



Supplementary Fig. S1. Variance stabilization at the narrative onset. The shaded region (Bins 0–1) represents the excluded warm-up period, with the analysis window commencing at Bin 2, where the signal enters a stable plateau.



Supplementary Fig. S2. Validation of surprisal-based trajectory clustering. Identification of optimal cluster count (k) using Elbow (Inertia) and Silhouette analysis. The vertical dashed line at k=5 denotes the point of optimal structural stability.

Supplementary Table S1. Metadata overview of the Korean flash fiction corpus.

(a) Metadata Schema Fields

Column Name	Description
isbn	International Standard Book Number
story_title	Title of the individual flash fiction story
text	Full narrative text of the story (Not provided due to copyright restrictions)
author	Name of the author
birth_year	Birth year of the author
gender	Gender of the author
book_title	Title of the source book or anthology
publisher	Name of the publisher
pub_year	Year of publication
country	Country of author

(b) Gender and Country Distribution

Category	Sub-category	Count	Percentage (%)
Gender	Male	1,200	41.1
	Female	1,055	36.1
	Unknown	666	22.8
Country	South Korea	2,666	91.3
	Japan	158	5.4
	USA	68	2.3
	Others*	29	1.0
Total		**2,921**	**100.0**

Note : Others include UK, France, Hungary, Russia, Canada, Spain, Ireland, and China.

(c) Gender and Origin Distribution

Metric	Value
Mean Year	2018.4
Median Year	2020.0
Standard deviation	4.9
Range (Min - Max)	1998 – 2025
Interquartile Range (IQR)	2016 – 2022

Supplementary Table S2. Hardware and software configuration of the high-performance computing (HPC) environment.

Component	Specification
Operating system	Ubuntu 24.04.2 LTS
CPU	Intel(R) Xeon(R) Silver 4516Y+
System memory (RAM)	251 GiB
GPU configuration	2 × NVIDIA RTX A6000
GPU memory (per device)	48 GB VRAM
Total GPU memory	96 GB VRAM (2 × 48 GB)

Note: These specifications represent the computational environment used for narrative signal extraction and all subsequent statistical analyses reported in this study.

Supplementary Table S3. Quantitative assessment of initialization noise and surprisal variance stabilization.

Bin Index	Global Variance	Variance Drop (%)
0	1.5555	-
1	1.1644	25.1%
2	0.8186	29.7%
3	0.7137	12.8%
4–9	0.71–0.78	< 4.0% fluctuation

Note : Summary of surprisal variance and step-wise reduction rates used to identify the stabilization threshold at the narrative onset. The significant drop in variance during the initial bins (0–1) reflects the dissipation of initialization noise, with stabilization (fluctuations < 4.0%) achieved from bin 4 onwards

Supplementary Table S4. Diagnostic tests for normality and homogeneity of variance across surprisal-trajectory shape descriptors.

Variable	Shapiro-Wilk (min p)	Levene's test (p)	Assumptions Met
Peak Position	< .001	< .001	No
Peak Intensity	< .001	< .001	No
Peak Count	< .001	.358	No
Volatility	< .001	.651	No

Note: Normality was assessed for each cluster using the Shapiro-Wilk test; the minimum -value across all clusters is reported (min p). Homogeneity of variance was assessed using Levene's test across all five clusters. An alpha level of 0.05 was used for all diagnostic tests. Since all variables significantly violated the assumption of normality (p <.001), non-parametric Kruskal-Wallis tests were employed for subsequent group-comparison analyses.

Supplementary Table S5. Pairwise post-hoc comparisons (Dunn's test) for surprisal-trajectory shape descriptors.

Comparison	Archetype 0	Archetype 1	Archetype 2	Archetype 3	Archetype 4
Peak Position					
Archetype 0	-	< .001	< .001	< .001	< .001
Archetype 1	< .001	-	< .001	< .001	< .001
Archetype 2	< .001	< .001	-	0.160	< .001
Archetype 3	< .001	< .001	0.160	-	< .001
Archetype 4	< .001	< .001	< .001	< .001	-
Peak Intensity					
Archetype 0	-	< .001	1.000	< .001	1.000
Archetype 1	< .001	-	< .001	< .001	< .001
Archetype 2	1.000	< .001	-	< .001	1.000
Archetype 3	< .001	< .001	< .001	-	< .001
Archetype 4	1.000	< .001	1.000	< .001	-
Peak Count					
Archetype 0	-	< .001	1.000	< .001	1.000
Archetype 1	< .001	-	< .001	0.232	< .001
Archetype 2	1.000	< .001	-	< .001	1.000
Archetype 3	< .001	0.232	< .001	-	< .001
Archetype 4	1.000	< .001	1.000	< .001	-

Note: p-values are adjusted using the Bonferroni method. Volatility is omitted as the omnibus test was non-significant.

Supplementary Table S6. Summary of representative short stories for the identified narrative archetypes.

Archetype	Rank	Author	Title [Translation]	Summary (Narrative Focus)	Peak Sentence [Translation]
0	1	Shin Ju-hee	Ko-ro-na-44 Geuk-bok-gi [Overcoming COVID-44]	A recovery journey from COVID-19 and psychological cleansing via cleaning a refrigerator.	Mundeuk, meomchugo sip-eotda. [Suddenly, I wanted to stop.]
	2	Kim Geun-pyeong	Chim-tu [Infiltration]	The tension of physical advances and strict boundaries in a cinema setting.	Geunyang jomul-ttag-ida. [It is merely a slight fondling.]
	3	Park Si-ul	Ut-eum Ju-sa [Laughter Injection]	A woman experiences uncontrollable laughter during a funeral after a medical "laughter injection."	Shin Daeri-cheoreom jjing-geurigo danineun hwasang-eun geumbang neul-eobeorindago. [A face that frowns like yours, Mr. Shin, will age quickly.]
1	1	Shin Kyung-sook	Ha-neun-im-ui Gu-du [God's Shoes]	An epistolary reflection on Van Gogh's life and the inherent healing power of art.	H-ga nabodaneun neoege hwolssin doumi doel mal-eul hae jul su isseul geot gat-aseoyeotji. [I thought H... could give you much more helpful advice than I could.]
	2	Jeong Seon-yeop	Mo-mo-nim [Momo-nim]	An aspiring writer faces hierarchy and naming politics within the literary establishment.	Eojjeon il-iya? Jugeun jul aratneunde? [What is the matter? I thought you were dead.]
	3	Lee Ki-ho	Eo-tteon Jol-eop-sik [A Certain Graduation]	A convenience store worker's attendance at a boy's graduation leads to class-based realization.	Pyeon-uijeom-eun gakkap-guna, urideul-eun jaran-da. [The convenience store is close, and we grow up.]
2	1	Jo Kyung-ran	Geum-yo-il [Friday]	A man's birthday dinner is derailed by the discovery of a hammer—his tool of community service—in his bag.	Junghakgyo gyosain ttal-eun... sangsehi gieokhago jijeokhaetda. [His daughter... pointed out in detail how much Song's lateness had worried and distressed the family.]
	2	Lee Ki-ho	Dal-ryeo-ra A-deul [Run, Son]	A father's dreams for his son's soccer talent collide with the reality of urban competition.	Geu-ui yeop jaseok-eneun... yeol sal meogeun adeul-i an-ja isseotda. [His ten-year-old son... was sitting in the seat next to him.]
	3	Kwak Seong-geun	Byeok-ryeon-hang Hwoe-jip [Raw Fish Restaurant at Byeokryeon Port]	A man planning an affair is surprised by his wife's proposal for a "gray divorce."	Corona ttaemune dadeul simgakhae. [Everyone is serious because of COVID.]

3	1	Gu Ja-myeong	Neo-wa Na-ui Ye-jeong-doen Gae-ul [Our Predestined Autumn]	A woman presentation of a contract-termination document marks the end of a 30-year marriage.	Kakisaek jeompeoga siljongja-ui silcheseong-eul jeung-geohamyeo nama isseotda. [A khaki jumper with a faded collar remained, attesting to the reality of the missing person.]
	2	Jeong I-su	Ri-mo-keon [Remote Control]	A wife's attempt to control her husband's alcoholism through diluted drinks and hidden medicine.	Chwichim sigan-i ahop si-eseo du sigan-ina dwi-ro mulleonatda. [His bedtime moved back by two hours from nine o'clock.]
	3	Lim Sang-tae	Jjal-beun Dong-haeng [Short Accompaniment]	A widower reflects on his brief life with his late wife through the imagery of a mountain motel.	Ae-cho-e yeohaeng-eun gwabun-han geos-iyeotda. [The trip was unwarranted in the first place.]
4	1	Kim Geun-pyeong	Kkae-da-reum [Realization]	A brief, celebratory account of South Korea's growth into a global cultural powerhouse.	Naga-boseyo. [Go outside.]
	2	Birdsmith	Ro-seu-teu Bi-seu-teu [Lost Beast]	A surreal urban narrative exploring digital isolation and the absurdity of modern social cues.	Chang bak haneul-e taeyang-i tteoolla bosong-bosong-han hayan gureum-e bul-eul noatda. [The sun rose in the window sky and set fire to the fluffy white clouds.]
	3	Yoo Kyung-sook	Jin-nun-kkae-bi-ro In-ha-yeo [Due to Sleet]	A young girl's path to becoming a nun is changed by a sleet storm, shifting her religious journey.	Seunim-i geu dongne-ro siju-reul naomyeon Suk-gyeong-ine sarangbang-eseo harubam-eul muk-eo gatda. [Whenever the nun came to beg, she would stay overnight in Suk-gyeong's guest room.]

Note : *Due to copyright restrictions governing contemporary literary works, the full text of the short stories cannot be publicly shared. To ensure academic transparency and allow for verification of the thematic clusters, we provide concise English summaries of the prototypical narratives (ranked by proximity to the archetype centroid). These summaries highlight the core plot points and character motivations that served as the qualitative basis for our computational modeling, while fully adhering to intellectual property laws.*

Supplementary Table S7. Point-wise diagnostic tests and omnibus group comparisons for discourse signals at surprisal peaks

Metric	Category	Factor/Archetype	Statistic	P-value
Coherence	Normality (Shapiro-Wilk)	Archetype 0	W: 0.9917	< 0.001
		Archetype 1	W: 0.9915	< 0.001
		Archetype 2	W: 0.9945	0.014
		Archetype 3	W: 0.9930	0.002
		Archetype 4	W: 0.9943	0.002
	Homogeneity (Levene)	All Clusters	F: 2.0940	0.079
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 7.7885**	**0.099**
Semantic Shift	Normality (Shapiro-Wilk)	Archetype 0	W: 0.9967	0.161
		Archetype 1	W: 0.9893	< 0.001
		Archetype 2	W: 0.9945	0.015
		Archetype 3	W: 0.9928	0.002
		Archetype 4	W: 0.9938	0.001
	Homogeneity (Levene)	All Clusters	F: 5.5143	< 0.001
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 13.6708**	**0.008**

Supplementary Table S8. Post-hoc pairwise comparisons (Dunn's test) for point-wise signal deviations at surprisal peaks.

	Archetype 0	Archetype 1	Archetype 2	Archetype 3	Archetype 4
Archetype 0	1.000	0.329	1.000	0.538	0.066
Archetype 1	0.329	1.000	0.241	1.000	1.000
Archetype 2	1.000	0.241	1.000	0.400	**0.044***
Archetype 3	0.538	1.000	0.400	1.000	1.000
Archetype 4	0.066	1.000	**0.044***	1.000	1.000

Note: Values represent Bonferroni-adjusted p-values.

Supplementary Table S9. Event-locked diagnostic tests and omnibus group comparisons for narrative recovery dynamics.

Metric	Category	Factor/Archetype	Statistic	P-value
Coherence TTR	Normality (Shapiro-Wilk)	Archetype 0	W: .923	< .001
		Archetype 1	W: .928	< .001
		Archetype 2	W: .934	< .001
		Archetype 3	W: .909	< .001
		Archetype 4	W: .927	< .001
	Homogeneity (Levene)	All Clusters	F :0.830	.506
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 12.063**	**.017**
Coherence Slope	Normality (Shapiro-Wilk)	Archetype 0	W: .940	< .001
		Archetype 1	W: .919	< .001
		Archetype 2	W: .911	< .001
		Archetype 3	W: .931	< .001
		Archetype 4	W: .942	< .001
	Homogeneity (Levene)	All Clusters	F :1.801	.126
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 7.169**	**.127**
Semantic Shift TTR	Normality (Shapiro-Wilk)	Archetype 0	W: .923	< .001
		Archetype 1	W: .943	< .001
		Archetype 2	W:.932	< .001
		Archetype 3	W: .932	< .001
		Archetype 4	W:.937	< .001
	Homogeneity (Levene)	All Clusters	F :0.838	.501
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 10.590**	**.032**
Semantic Shift Slope	Normality (Shapiro-Wilk)	Archetype 0	W: .910	< .001
		Archetype 1	W:.924	< .001
		Archetype 2	W: .923	< .001
		Archetype 3	W: .939	< .001
		Archetype 4	W:.937	< .001
	Homogeneity (Levene)	All Clusters	F : 0.425	.791
	Omnibus (Kruskal)	**Inter-cluster Diff.**	**H: 13.077**	**.011**

Note: All p-values for the Shapiro-Wilk test were below the .001 threshold, indicating significant deviation from normality across all archetypes and metrics. Consequently, non-parametric Kruskal-Wallis tests were employed for inter-cluster comparisons.

Supplementary Table S10. Post-hoc pairwise comparisons (Dunn's test) for event-locked narrative recovery metrics across archetypes.

Comparison	Archetype 0	Archetype 1	Archetype 2	Archetype 3	Archetype 4
Coherence TTR					
Archetype 0	-	1.000	1.000	0.349	1.000
Archetype 1	1.000	-	1.000	**0.016***	1.000
Archetype 2	1.000	1.000	-	**0.045***	1.000
Archetype 3	0.349	**0.016***	**0.045***	-	0.392
Archetype 4	1.000	1.000	1.000	0.392	-
Semantic Shift TTR					
Archetype 0	-	0.091	0.371	1.000	1.000
Archetype 1	0.091	-	1.000	0.153	0.822
Archetype 2	0.371	1.000	-	0.567	1.000
Archetype 3	1.000	0.153	0.567	-	1.000
Archetype 4	1.000	0.822	1.000	1.000	-
Semantic Shift Slope					
Archetype 0	-	1.000	1.000	**0.040***	1.000
Archetype 1	1.000	-	1.000	0.331	1.000
Archetype 2	1.000	1.000	-	**0.010***	0.836
Archetype 3	**0.040***	0.331	**0.010***	-	0.626
Archetype 4	1.000	1.000	0.836	0.626	-

Note: Values represent Bonferroni-adjusted p-values. Cohrece Slope is omitted as the omnibus test was non-significant.