

Structural Patterns in Award-Winning Novellas

Milka Kaplan^{1,*†}, Armin Shmilovici^{1,†} and Mark Last^{1,†}

¹Faculty of Computer and Information Science, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva, Israel

Abstract

The long-term goal of this research is to evaluate the accuracy of known literature patterns in narrated text. Following Lima et al. who suggested literary patterns for each of the following five narrative genres: *Comedy*, *Mystery*, *Romance*, *Satire* and *Tragedy*, we explore the prevalence of those patterns (or their elements) are indeed prevalent in Award-Winning Novellas.

We have manually annotated 35 award-winning novellas for their narrative genres and compared the results to automated annotation by a Large Language Model (LLM), which analyzes the five narrative genre patterns and their structural elements. Statistical tests (such as binomial test, Herfindahl-Hirschman index (HHI)) indicate that there is a high recall and low precision of the LLM with respect to human annotation.

Our feature selection analysis reveals that *Closure*, *Final Confrontation*, and *Return* are the most structurally central pattern elements across all genres. We also perform an element by element analysis to identify the structural elements, which are most informative for genre identification. This study provides the first step towards our long-term goal by proposing a methodology evaluation of literature patterns and demonstrating the proposed methodology on five common patterns in a small dataset of 35 award-winning novellas.

Keywords

Automatic Literary Critic, Novellas, Computational Narratology, Narrative Structure, Genre Analysis, Literary Structure Evaluation

1. Introduction

The purpose of this article, which relates to the field of computational narratology, is to explore whether the currently known narrative structures can represent award winning narrated texts.

One of the common narrative structures is Joseph Campbell's Hero's Journey "*Hero's Journey*" [1], presented in *The Hero with a Thousand Faces* (2004, original work published 1949). Campbell outlined the narrative structure as a cyclical twelve-stage journey involving departure, initiation, and return. His model has been applied to both literary epics and modern cinema. While Campbell does not offer quantitative metrics, evaluation of this model often considers the completeness of the cycle, the protagonist's transformation, and the emotional response elicited in the audience. Campbell's *The Structure of Monomyth* effectively complemented by Christopher Vogler's book *A Writer's Travel Guide for Future Authors* "A *Writer's Travel Guide for Future Authors*" [2]. Inspired by Campbell's suggestion to J. Vogler, we consider literary narrative patterns as a way to create an interactive composition at a thematically consistent plot level.

Based on this theoretical perspective we expect that those literary narrative patterns will be prevalent in the writing of award-winning authors, unless, they are deficient or ill defined. In particular, we consider only the five main narrative genres as defined by Lima et al. (the PatternTeller [3]).

There are two novel aspects in this research: a) The potential use of LLM for large scale literary analysis; b) Investigating the usefulness of five of the narrative patterns in PatternTeller [3] for analyzing award winning Novellas. This research method can be extended to exploring the prevalence of many types of literary patterns in narrated texts.

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): *Proceedings of the Text2Story'26 Workshop, Delft (The Netherlands), 29-March-2026*

*Corresponding author.

† These authors contributed equally.

✉ milkak@post.bgu.ac.il (M. Kaplan); armin@bgu.ac.il (A. Shmilovici); mlast@bgu.ac.il (M. Last)

🌐 https://github.com/MilkaKaplan/Narrative_Pattern_Evaluation (M. Kaplan)

🆔 0009-0002-7689-8606 (M. Kaplan); 0000-0002-7202-6360 (A. Shmilovici); 0000-0003-0748-7918 (M. Last)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The rest of the paper is organized as follows: Section 2 presents a short survey of related works; Section 3 describes the proposed structural analysis methodology; Section 4 presents our experimental results; Section 5 evaluates the importance of various narrative elements; and Section 6 concludes with future research directions.

2. Related Work

Frye’s archetypal criticism, presented in *Anatomy of Criticism* [4], establishes four fundamental narrative modes – comedy, romance, tragedy, and irony/satire - associated metaphorically with the seasons of the year. Comedy corresponds to spring, representing the optimistic view of youth and conformity to a just world. Romance embodies summer, marking the call to heroic action to restore disruptions in world order. Tragedy reflects autumn, where pride (hubris) invites punishment (nemesis) from unforgiving powers. Satire represents winter, evoking defeat and impotence in an unjust world that can only be criticized through irony. While epic and dramatic plots trace back to Aristotle, epistemic plots emerged only in the nineteenth century with the detective fiction genre.

Ryan [5] offers a complementary typology distinguishing three plot types for interactive narratives: epic plots (physical actions, epitomized by Campbell’s Monomyth), dramatic plots (networks of human relations, comprising tragedy and comedy), and epistemic plots (driven by the desire to know, characteristic of mystery narratives).

Lima et al. [3] proposed PatternTeller, a system for AI-powered story composition guided by genre patterns. They defined the Most Specific Generalization (MSG) method and 11 different literary patterns for generating stories using LLM. Combining Ryan’s and Frye’s proposals, they identify five fundamental genres: Comedy, Romance (in the sense of epic plots), Tragedy, Satire, and Mystery. *In this research we implement LLM as a tool for literary text analysis using their specific definitions for literary patterns* to implement LLM as a tool for literary text analysis. For example, their Tragedy pattern comprises 7 stages: (1) Stage of Normalcy; (2) Catalyst – a disruptive event; (3) Awakening – protagonist questions their values; (4) Confrontation – internal/external conflicts; (5) Crisis – pivotal decision; (6) Downfall – loss or fall from grace; (7) Final Act – consequences and resolution [3]

3. Evaluation Methodology

The above definition of literary patterns raises the question of whether those patterns can effectively distinguish between genres of novellas? Are those structural patterns useful in distinguishing a group of texts of one genre from another? Or are they reduced to a universal narrative pattern?

For our preliminary study we assembled a corpus of 35 highly rated English novellas [6] (7,500 to 102,810 words long) grouped into 5 genres (Tragedy, Satire, Romance, Mystery and Comedy) in from different periods, e.g. Oscar Wilde’s “*The Canterville Ghost*” from 1887, William Golding’s “*Lord of the Flies*” from 1954 and Courtney Milan’s “*The Governess Affair*” from 2012 (Table 4 in Appendix A). A human expert (the first author) evaluated the compliance of every novella with each of the five genres on a scale from 0 to 4 points, where “0” – no correspondence, “1” – weak correspondence, “2” – moderate correspondence, “3” – strong correspondence and “4” – very strong correspondence.

Regarding the structural features (chapters or paragraphs) in the texts, chapter titles, section headings, and structural markers (e.g., “Part 1”, “Chapter XX”, page numbers) were explicitly removed from all texts prior to annotation by both human annotators and the LLM. This decision was motivated by the need to prevent *metadata leakage*: explicit structural markers can act as strong priors that artificially guide stage assignment - for instance, a label like “Part III: The Downfall” would trivially suggest a late-narrative stage, bypassing the actual semantic reasoning we aimed to evaluate. By stripping these markers, we ensured that both human and LLM annotations reflect genuine narrative comprehension rather than surface-level structural cues. We acknowledge that this preprocessing choice introduces a methodological trade-off: while it strengthens the validity of the annotation process, it diverges from the natural reading experience and may affect how chunking boundaries are perceived. In our future

experiments, we plan to rerun the annotation on original, unstripped texts would be a valuable future experiment to assess the magnitude of this effect.

To assess the extent to which the MSG-derived Campbell’s patterns extracted via MSG [3] (hereafter referred also to as PatternTeller) can accurately classify the genre of each novella from the same corpus, we developed a systematic scoring protocol for LLM-based analysis. The *Claude haiku-3.5* model was chosen for evaluating the text of each novella due to the large size of its context window (approximately 150,000 words). Each Campbell’s pattern consists of a fixed number of elements (e.g., 8 elements for Satire, 10 elements for Romance). The protocol employs a bottom-up aggregation approach, where individual structural elements are evaluated first, then combined into an overall genre correspondence score. For each element, the LLM-based evaluator is prompted to assigns an element score on a three-point scale: “0” – The element is absent in the narrative; “1” – The element is partially developed, weak, or implicit; “2” – The element is clear, strong, and explicit.

The Total Stage Score (S) is computed as the sum of all individual element scores. *Coverage* is then calculated as the ratio of the obtained score to the maximum possible score:

$$\text{Coverage} = \frac{\sum_{i=1}^{N_{\text{stages}}} s_i}{2 \cdot N_{\text{stages}}}$$

where s_i is the result of the element i , N_{stages} - the number of elements in the narrative pattern. For example, in the 8-element model, the maximum score is 16, so a total score of 12 gives a coverage score of 0.75.

The Coverage value is converted to a Final Score on a 0-4 scale using predefined thresholds-based conversion to ensure consistent scoring across narrative patterns with different number of elements and provides a clear interpretability: a score of 4 indicates strong correspondence (86 scores and more - clearly present), while a score of 0 indicates minimal correspondence. For example, in the novella *Portrait of Dorian Gray*, 14 points out of 16 were obtained for all stages of Campbell’s pattern. Coverage was obtained with a ratio of $14/16 = 0.875$. which is the threshold for awarding a grade of 4 in the whole pattern (Figure 2 in Appendixes).

Simultaneously with assigning numerical scores, each component of the stage-based pattern assessment is accompanied by a brief explanation, and a detailed textual justification of the final score is also generated. A full example the “Tragedy” pattern of the model’s response is provided in Figure 1 in Appendix B.

This methodology was applied to all five genre patterns (Comedy, Mystery, Romance, Satire, Tragedy) across the 35 novellas in the corpus, yielding a 35×5 matrix of genre correspondence scores for comparison with human expert assessments.

4. Result and Analysis

The following hypotheses were formulated in order to test whether it is possible to classify novellas by genre according to Campbell’s patterns. Since a narrative text may belong to more than one genre (e.g. a romantic comedy), we tested the following hypotheses on the two most likely genres of each novella:

H0: The agreement between the Human and PatternTeller’s selection of the top-2 genres is not better than a random chance.

H1: Human and PatternTeller are likely to agree on the two best genres of each novel (out of 5).

The choice of top-2 genres reflects the multi-generic nature of novellas which frequently blend two different genre conventions (e.g. comedy and satira).

As an example, we consider William Golding’s “*Lord of the Flies*” (1954). Stranded on an uninhabited island, a group of British schoolboys attempt self-governance under Ralph’s leadership. The tragedy unfolds through three key figures: Ralph (civilization), Jack (savagery), and Piggy – whose rationality proves powerless against violence, and whose death symbolizes the destruction and savagery.

We see the following score estimates for the 7 elements of the tragedy template unfold in “*Lord of the Flies*”: *Normalcy*: 2 (Ralph and the boys create initial social structure); *Catalyst*: 2 (Fear of the “beast”);

Awakening: 2 (“Maybe there is a beast... maybe it’s only us” - Simon); *Confrontation*: 2 (Conflicts between Ralph and Jack, civilized vs. savage impulses); *Crisis*: 2 (Simon, is mistaken for a beast and killed); *Downfall*: 2 (Moral collapse, tribal violence); *Final Act*: 2 (the island is on fire, rescue has arrived, but too late). The Coverage is 14/14 and the aggregated Tragedy pattern score reached its maximum value of 4.

4.1. Binomial test

To compare the consistency between the human and the LLM evaluations of Campbell narrative patterns (PatternTeller), we labeled each novella with genre having a coverage score 3 or higher. This binarization revealed a clear asymmetry: human expert’s estimates provided an average of 1.37 genres per novella, while the LLM based on PatternTeller provided an average of 2.83 genres. For example, for “*Lord of the flies*” the human ranked Satire (score 3) and Tragedy (score 4) as top coverage genres, while the LLM ranked Comedy (score 4), Mystery (score 3), Romance (score 3), Satire (score 4), Tragedy (score 4). (The LLM considered that Piggy’s pedantic complaints create a comic atmosphere).

For each novella, a match is considered between the human and LLM scores if they match by at least one common genre of the top 2 genres. In case of ties, genres were selected in alphabetical order (Comedy, Mystery, Romance, Satire, Tragedy). There were 25 matches out of the 35 novellas, whereas the expected number of random matches is 24.5, resulting in a p-value of 0.51. Although the scoring consistency turned out to be almost equal to a random and it is not enough to consider it statistically significant for this relatively small corpus of novellas. Meaning that agreement may partly reflect the LLM’s tendency to subsume the human’s choice within a broader set of genre assignments rather than genuine convergence.

Notably, when the human annotator also operated within the PatternTeller framework, the agreement with the LLM based on PatternTeller did not improve despite the average number of genres for human based on PatternTeller being 1.83 which is closer to the value for LLM based on PatternTeller (2.83). This suggests that PatternTeller rather introduces systematic divergence: both annotators assign several genres, but different ones.

4.2. Herfindahl–Hirschman Index

The Herfindahl–Hirschman Index (HHI) [7, 8] allowed one to determine the concentration of genres within a single novella.

$$HHI = \sum_{i=1}^N \left(\frac{s_i}{\sum_{j=1}^N s_j} \right)^2$$

where s_i is the score assigned to an element of the genre i , and $N = 5$ is the number of genres. The Effective Number of Genres is calculated as $N_{\text{eff}} = \frac{1}{HHI}$.

Human annotations based on general genre knowledge show a significantly higher genre concentration ($HHI = 0.37$, $N_{\text{eff}} = 2.64$), reflecting clear and discriminative genre judgements. When the human annotator operated within the PatternTeller framework, concentration decreased substantially ($HHI = 0.26$, $N_{\text{eff}} = 3.83$, $p < 0.001$). The LLM based on PatternTeller shows the lowest concentration ($HHI = 0.21$, $N_{\text{eff}} = 4.72$), essentially responding “yes” to all genre classifications simultaneously. *This behavior indicates that PatternTeller’s patterns are insufficiently distinct to function as discriminative classifiers.* Notably, among all LLMs tested, Gemini most closely approximates human PatternTeller-based genre concentration ($N_{\text{eff}} = 3.62$ vs. 3.83), suggesting it is somewhat more selective in applying PatternTeller patterns than the other models.

An effective genre classification system should assign significantly higher scores to texts within a genre than to texts outside that genre. The results of statistical tests indicate that the extracted patterns reflect universal narrative structures rather than genre-specific features.

To assess the extent to which a LLM (Claude) reproduces human genre judgments, we computed classification metrics using a threshold ≥ 3 as positive. This value corresponds to a “strong fit” on our

0-4 scoring scale, ensuring that only genres with a clear and well-supported structural presence are treated as positive matches. Scores below this threshold indicate partial or ambiguous alignment and hence they excluded to reduce noise in the classification.

The results in Table 1 indicate moderately strong agreement between human and LLM annotations (Precision = 0.78, Recall = 0.75, F1 = 0.74). However, the LLM does not perfectly replicate human judgments, revealing systematic discrepancies across genres. The same metrics between the Human vs PatternTeller: Precision = 0.33, Recall = 0.89, F1 = 0.47. The results indicate that PatternTeller rarely misses genres identified by the Human, but tends to assign an excessive number of genre labels, prioritizing recall over precision.

Table 1

Genre-wise agreement metrics for Human-LLM and Human-PatternTeller comparisons

Genre	Human-LLM *Claude							Human (PatternTeller)-LLM (PatternTeller) *Claude						
	TP	FP	FN	TN	Precision	Recall	F1	TP	FP	FN	TN	Precision	Recall	F1
Comedy	3	5	2	15	0.38	0.60	0.46	23	9	2	1	0.72	0.92	0.81
Romance	6	2	0	17	0.75	1.00	0.86	0	18	0	17	0.00	0.00	0.00
Tragedy	6	0	2	17	1.00	0.75	0.86	21	14	0	0	0.60	1.00	0.75
Satire	6	2	1	16	0.75	0.86	0.80	13	16	1	5	0.45	0.93	0.60
Mystery	6	0	5	14	1.00	0.55	0.71	8	12	0	15	0.40	1.00	0.57
TOTAL / AVG	27	9	10	79	0.78	0.75	0.74	65	69	3	38	0.43	0.77	0.55

Comedy exhibits the lowest precision (0.38) despite moderate recall (0.60), indicating that the LLM frequently identifies comedic elements in texts where the human expert does not. *This is consistent with the known ambiguity of comedy as a genre and suggests that comedic elements are broadly distributed across narratives even when not central to the work.* Romance achieves perfect recall (1.00) - the LLM identifies all human-labeled romance texts - though with some false positives (Precision = 0.75). Tragedy demonstrates the highest F1 score (0.86), indicating strong human-LLM agreement for this genre.

We additionally tested the PatternTeller template on other LLMs (Table 2).

Table 2

Genre-wise agreement metrics for Human (PatternTeller)-LLM (PatternTeller) comparisons

Genre	Human (PatternTeller)-LLM (PatternTeller) *Gemini							Human (PatternTeller)-LLM (PatternTeller) *DeepSeek						
	TP	FP	FN	TN	Precision	Recall	F1	TP	FP	FN	TN	Precision	Recall	F1
Comedy	15	6	10	4	0.71	0.60	0.65	14	6	11	4	0.70	0.56	0.62
Romance	0	24	0	11	0.00	0.00	0.00	0	1	0	34	0.00	0.00	0.00
Tragedy	14	1	7	13	0.93	0.67	0.78	8	3	13	11	0.73	0.38	0.50
Satire	10	4	4	17	0.71	0.71	0.71	13	12	1	9	0.52	0.93	0.67
Mystery	5	5	3	22	0.50	0.63	0.56	5	5	3	22	0.50	0.63	0.56
TOTAL / AVG	44	40	24	67	0.57	0.52	0.54	40	27	28	80	0.49	0.50	0.47

The low agreement between human and LLM based on PatternTeller (F1=0.47-0.54) contrasts with the agreement between human and LLM based on the basic genre classification (F1=0.74). To investigate whether this discrepancy could be attributed to inconsistency in LLM annotation under the PatternTeller framework, we conducted an inter-LLM comparison in which three LLMs (Claude, Gemini, and DeepSeek) independently annotated the same texts using PatternTeller’s structural patterns (Table 3). If the LLM were consistent in their interpretation of PatternTeller’s framework, we would expect high mutual agreement among them.

However, the level of agreement between LLMs based on PatternTeller remained low (F1=0.57-0.61 in Table 3), which is comparable to the level of agreement between LLM and a human in the same condition (F1=0.47-0.54 in Table 2). This result, in fact, casts doubt on the assumption that the reason for the observed discrepancies may be the human factor.

Table 3

Agreement metrics for inter-LLM comparisons under the PatternTeller framework

Genre	LLM (PatternTeller) * Claude – LLM (PatternTeller) * Gemini							LLM (PatternTeller) * Claude – LLM (PatternTeller) * DeepSeek						
	TP	FP	FN	TN	Precision	Recall	F1	TP	FP	FN	TN	Precision	Recall	F1
Comedy	20	1	12	2	0.95	0.63	0.75	19	1	13	2	0.95	0.59	0.73
Romance	16	8	2	9	0.67	0.89	0.76	1	0	17	17	1.00	0.06	0.11
Tragedy	15	0	20	0	1.00	0.43	0.60	11	0	24	0	1.00	0.31	0.48
Satire	13	1	16	5	0.93	0.45	0.60	23	2	6	4	0.92	0.79	0.85
Mystery	5	5	15	10	0.50	0.25	0.33	10	0	10	15	1.00	0.50	0.67
TOTAL / AVG	69	15	65	26	0.81	0.53	0.61	64	3	70	38	0.97	0.45	0.57

5. Selection of Key Narrative Elements for Each Genre

There is a total of 41 different structural elements in the five different genre models that were estimated for each of the 35 novellas. A close inspection of their description discovers resemblance between some of them. Therefore, we would like to identify which structural elements are most unique (therefore, informative) in each genre, and to identify which narrative elements are most central to the shared narrative space across different genres and whose removal would most strongly disrupt overall narrative coherence which structural elements are most similar across different genres.

We applied a multi-task L1/L2 regularized regression (MultiTaskLasso) in a self-reconstructive setting[9] on the 35×41 matrix after normalizing it.

This approach identifies features that are most informative for explaining the joint variation across the entire corpus, i.e. narrative elements that serve as structural anchors independent of genre-specific realization.

The three highest-ranking elements discovered were *Closure*, *Final Confrontation* and *Return* (Table 5 in Appendix). Taken together, the results suggest that despite genre diversity, the analyzed novellas are dominantly organized around classical dramatic arcs featuring a clearly articulated crisis, confrontation/insight, and eventual return to a transformed order. We also detected the 1-3 most unique structural elements in each genre, such as *Uncovering secrets and clues* - the 4th structural element in a mystery or *Downfall – loss or fall from grace* - the 6th structural element in a tragedy (Table 6 in Appendix).

6. Conclusions and Future Research

The purpose of this article is to explore the prevalence of five common narrative structure models in award winning novellas, using LLM as an annotator assistant. The corpus of 35 novellas was specifically designed as a pilot study to test the methodology before scaling. The methodology presented here can be used to evaluate the strength of those theoretical narrative structure models for analyzing literature. From the preliminary results on a small corpus, we can draw the following conclusions:

1) An LLM can aid in annotating complex narrative patterns, yet it must be tuned and evaluated by human experts, since its annotations may not be sufficiently accurate.

2) PatternTeller’s structural patterns introduce systematic interpretive ambiguity. Agreement between human and LLM annotations under the PatternTeller (F1=0.47-0.54) was substantially lower than direct Human-LLM agreement, and inter-LLM agreement under the same framework (F1=0.57-0.61) was similarly low. HHI analysis reveals that LLM (Claude) using PatternTeller assigns an effective number of 4.72 genres per text, essentially responding “yes” to all genre classifications simultaneously. However, Gemini most closely approximates human (based on PatternTeller) genre concentration.

3) There are many structural communalities in different narrative patterns that are fairly genre independent. Despite genre diversity, the analyzed novellas are dominantly organized around resolution-phase structural elements: *Closure*, *Final Confrontation*, and *Return* emerged as the most structurally central elements across all genres, suggesting that PatternTeller-derived patterns prioritize narrative closure over genre-specific differentiation.

In future research, we will extend the corpus to the 600 novellas of [6]. This will significantly

increase the statistical power and allow us to draw more reliable conclusions; hire additional human annotators and calculate the agreement between them (Cohen's Kappa or Krippendorff's alpha); extend our LLM-assisted analysis to additional narrative patterns described in [3] and elsewhere. We will use the extended corpus to evaluate our hypothesis that LLMs can use valid structural patterns to annotate and analyze literature works.

Acknowledgments

This research was partially supported by the Data Science Research Center, Ben-Gurion University of the Negev, Ministry of Aliyah and Integration, Israel.

Declaration on Generative AI

The paper was written manually.

During the preparation of this work, the authors used Claude Haiku 3.5 (via Anthropic API) in order to: Generate experimental datasets for analysis. The authors reviewed all generated content, performed all analysis, drew all conclusions, and assume full responsibility for the publication's content.

References

- [1] J. Campbell, *The Hero with a Thousand Faces*, Princeton University Press, 2004.
- [2] C. Vogler, *The Writer's Journey: Mythic Structure for Writers*, Michael Wiese Productions, Studio City, CA, 2007.
- [3] E. S. de Lima, M. M. E. Neggers, A. L. Furtado, Multigenre ai-powered story composition, 2025. URL: <https://arxiv.org/abs/2405.06685>. `arXiv:2405.06685`.
- [4] N. Frye, *Anatomy of Criticism*, Princeton University Press, Princeton, NJ, 1957.
- [5] M.-L. Ryan, *Interactive narratives and plot types*, *New Media & Society* (2008).
- [6] World's greatest novellas, 2023. URL: https://www.goodreads.com/list/show/1236.World_s_Greatest_Novellas, goodreads list of highly rated novellas.
- [7] S. A. Rhoades, The herfindahl-hirschman index, *Federal Reserve Bulletin* 79 (1993) 188.
- [8] M. A. Adelman, Comment on the "h" concentration measure as a numbers-equivalent, *The Review of Economics and Statistics* (1969) 99–101.
- [9] N. Meinshausen, P. Bühlmann, Stability selection, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72 (2010) 417–473.

A. Corpus of Highly Rated Novellas

Table 4: Corpus of highly rated novellas selected from the Goodreads list [6].

Genre	Corpus filename	Year	Author	Word count	Avg. rating	Goodreads score
tragedy	002_mice-and-men.txt	1937	John Steinbeck	29673	3.89	31023
	302_the_great_gatsby.txt	1925	F. Scott Fitzgerald	48447	3.93	168
	000_the_old_man_and_the_sea.txt	1952	Ernest Hemingway	26269	3.81	26163
	001_animal-farm.txt	1945	George Orwell	29908	3.99	51254
	051_the_picture_of_dorian_gray.txt	1891	Oscar Wilde	54348	4.13	2753
	028_Lord_of_the_flies.txt	1954	William Golding	59725	3.70	4283
	093_to_kill_a_mocking_bird.txt	1960	Harper Lee	99235	4.26	1034
satire	067_the_catcher_in_the_rye.txt	1951	J.D. Salinger	74190	3.80	1684
	102_Miss_Lonelyhearts.txt	1933	Nathanael West	23203	3.75	941
	109_being_there.txt	1970	Jerzy Kosiński	25149	3.86	842
	156_the_13_clocks.txt	1950	James Thurber	11750	4.00	457
	258_farmer_giles_of_ham.txt	1949	J.R.R. Tolkien	16840	3.88	209
	317_the_wonderful_o.txt	1957	James Thurber	10757	3.87	152
	407_the_restaurant_at_the_end.txt	1980	Douglas Adams	52256	4.20	100
romance	055_giovannis_room.txt	1956	James Baldwin	54693	4.33	2647
	151_on_chesil_beach.txt	2007	Ian McEwan	36743	3.61	475
	270_the_governess_affair.txt	2012	Courtney Milan	32960	3.85	195
	377_greater_than_rubies.txt	2012	Hallee Bridgeman	56920	4.04	100
	418_breathless.txt	2012	Sophie Jordan	25116	3.80	99
	489_skin_deep.txt	2012	Pamela Clare	52170	4.12	94
	560_how_to_marry_a_werewolf.txt	2018	Gail Carriger	47391	4.19	81
mystery	054_doyle-hound-383.txt	1902	Arthur Conan Doyle	59321	4.13	2678
	135_the_maltese_falcon.txt	1930	Dashiell Hammett	66479	3.87	553
	036_the_turn_of_the_screw.txt	1898	Henry James	39471	3.39	3488
	191_every_heart_a_doorway.txt	2016	Seanán McGuire	38579	3.72	367
	337_the_library_window.txt	1896	Mrs. Oliphant	18403	3.71	137
	345_lyras_oxford.txt	2003	Philip Pullman	7354	3.62	132
	495_hammers_on_bone.txt	2016	Cassandra Khaw	17662	3.62	93
comedy	043_the_prime_of_miss_jean_brodie.txt	1961	Muriel Spark	39498	3.70	3121
	079_shopgirl.txt	2000	Steve Martin	36583	3.42	1316
	316_the_canterville_ghost.txt	1887	Oscar Wilde	11373	3.88	154
	046_cannery_row.txt	1945	John Steinbeck	49837	4.06	2958
	268_under_the_greenwood_tree.txt	1872	Thomas Hardy	56707	3.65	196
	452_the_perilous_life_of_jade.txt	2012	Zen Cho	24874	3.79	97
	201_the_tiny_wife.txt	2010	Andrew Kaufman	14313	3.61	343

B. Sample LLM Analysis Output

Below is a representative example of the raw output produced by the analysis pipeline for *Animal Farm* (Orwell, 1945).

```
FILE: Copy of Copy of 1_animal-farm.txt
PROCESSED: 2026-01-14 12:03:26
WORDS: 30,050
=====

I'll analyze this text using the Tragedy pattern.

=== MODEL EVALUATIONS ===

Model: Tragedy
StageTotal: 12/14
Coverage: 0.86
Score: 4
Justification: The text perfectly follows the tragic arc of a revolution betrayed, where
the protagonist (the animal collective) experiences a complete moral and social downfall
from their initial idealistic hopes.

Stage Breakdown:
1. Normalcy (2/2): Animals live under Jones's oppressive farm regime
2. Catalyst (2/2): Old Major's revolutionary speech inspires rebellion
3. Awakening (2/2): Animals overthrow Jones and establish Animal Farm
4. Confrontation (2/2): Internal conflicts between Snowball and Napoleon emerge
5. Crisis (2/2): Napoleon consolidates power through manipulation and violence
6. Downfall (1/2): Animals gradually lose their revolutionary principles
7. Final Act (1/2): Pigs become indistinguishable from humans, completing the tragic
transformation

The text demonstrates a comprehensive tragic structure where the initial noble ideals are
systematically corrupted, leading to a complete betrayal of the original revolutionary
vision. The animals' dream of equality is ultimately destroyed by the very leaders who
promised liberation.
```

Figure 1: Example of LLM analysis pattern output Tragedy for *Animal Farm*.

C. Full Annotation Scores

Name of novel	Comed				Myster				Myster				Roman				Satire				Tragedy				Comed				Roman				Tragedy				Satire				Myster			
	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an	Hum an									
Of Mice and Men	1	0	0	2	1	1	0	0	1	2	1	1	0	0	2	2	1	2	2	4	4	4	4	4	4	4	2	2	1	1	4	3	3	1										
The Great Gatsby	0	0	3	2	2	0	0	2	2	4	2	3	3	2	2	1	2	2	3	4	4	3	4	4	4	4	2	1	1	3	3	2	1											
The Old Man and the Sea	0	0	0	3	1	1	0	0	2	0	1	0	0	4	4	2	0	0	2	0	1	4	2	4	4	3	1	1	2	0	0	0	0											
Animal Farm	0	0	4	2	2	2	1	3	3	0	1	0	0	3	2	2	4	4	4	4	4	3	3	4	4	2	3	1	4	4	4	0	0											
The Picture of Dorian Gray	0	0	0	3	4	3	3	2	4	1	1	2	1	2	3	1	2	3	4	0	4	4	4	4	4	4	3	3	1	4	4	2	1											
Lord of the Flies	1	1	1	4	4	4	0	1	2	3	1	1	0	0	3	3	1	3	3	4	3	4	4	4	4	4	2	4	1	4	4	4	2											
To Kill a Mockingbird	0	0	1	3	3	3	1	2	2	1	1	4	1	3	4	4	3	3	2	3	3	4	4	4	4	4	3	4	2	4	4	3	1											
The Catcher in the Rye	0	0	0	3	1	3	0	2	2	1	1	1	1	2	2	1	1	3	3	4	3	2	2	2	4	3	2	4	2	4	4	4	1											
Miss Lonelyhearts	0	0	0	2	1	1	0	1	2	2	3	2	2	1	2	2	1	2	4	4	4	4	3	4	4	4	4	4	1	4	4	1	0											
Being There	2	3	3	4	3	0	2	1	1	1	1	1	1	2	2	1	4	4	3	0	2	0	1	3	2	2	3	1	2	2	2	0	0											
The 13 Clocks	3	3	3	4	4	4	3	1	4	1	3	2	3	4	4	4	4	1	2	4	1	4	1	0	4	1	2	2	1	2	2	2	1											
Farmer Giles of Ham	2	3	4	4	4	4	1	1	3	0	3	0	0	0	4	4	2	2	2	4	0	3	0	0	4	1	1	2	1	2	1	2	1											
The Wonderful O	1	4	4	4	3	0	2	3	1	2	1	2	1	3	4	1	3	4	4	3	4	0	4	1	1	3	1	3	1	2	2	1	0											
The Restaurant at the End of the Universe	1	4	4	4	2	3	3	3	1	2	0	1	3	3	1	3	4	4	4	1	4	1	1	1	4	1	2	2	0	2	2	2	0											
Giovanni's Room	0	0	0	3	1	2	0	1	2	4	2	4	2	2	2	1	1	1	4	4	4	4	4	4	4	4	4	4	1	4	4	4	1											
On Chesil Beach	0	0	0	3	1	2	0	1	2	3	1	2	1	2	1	2	1	1	1	4	4	4	4	4	4	4	4	4	0	0	4	4	2											
The Governess Affair	1	2	4	4	4	2	4	4	2	4	4	4	4	4	3	1	0	3	1	3	0	4	1	1	4	4	4	3	1	2	3	1	2											
Greater Than Rubies	0	0	2	4	4	4	0	1	4	1	1	3	4	2	3	1	0	0	2	0	0	2	1	4	4	1	2	1	2	1	2	1	0											
Breathless	1	1	3	2	3	3	1	2	1	1	2	4	4	3	1	1	0	3	0	0	2	2	1	4	4	2	1	2	1	2	0	1												
Skin Deep	1	1	1	4	4	4	0	2	4	1	4	3	4	4	4	2	1	0	3	0	0	1	1	4	1	2	2	1	2	1	2	1	0											
How to Marry a Werewolf	1	2	4	4	4	2	1	3	1	4	4	4	3	3	1	1	1	4	1	4	1	2	4	4	1	3	2	1	2	2	2	1	0											
The Hound of the Baskervilles	1	0	0	3	1	2	4	4	4	4	1	1	3	3	1	1	1	0	4	0	2	1	2	4	1	2	4	1	4	4	2	4	0											
The Maltese Falcon	0	0	0	3	1	2	4	4	3	4	4	1	1	2	3	1	0	2	4	4	1	1	3	4	2	3	4	1	4	4	4	4	0											
The Turn of the Screw	0	0	0	3	1	1	4	4	3	2	2	0	1	2	2	1	0	0	3	1	4	4	4	4	4	2	4	1	4	4	3	4	0											
The Perilous Life of Jade Yeo	1	1	1	3	2	2	4	4	4	4	2	1	4	4	3	1	0	2	3	0	4	3	2	3	4	3	2	3	1	3	3	3	4											
The Library	1	0	0	3	2	3	4	4	3	2	1	1	2	3	1	0	0	3	0	0	4	1	2	2	4	2	3	1	4	1	4	1	3											
Lyrá's Oxford	1	1	3	4	3	4	4	3	4	1	0	0	3	3	1	0	0	1	3	0	2	1	4	4	1	1	3	1	3	3	2	3	0											
Hammers on Bone	0	0	0	3	2	2	4	4	4	4	0	0	2	3	2	1	1	1	3	3	4	4	2	4	3	3	3	2	4	4	4	4	4											
The Prime of Miss Jean Brodie	1	1	1	3	2	3	1	2	2	4	1	1	2	2	3	2	1	2	3	4	4	4	2	4	4	4	2	3	1	3	3	1	1											
Shogirl	2	2	2	3	4	0	0	2	2	1	1	3	4	3	3	1	1	1	3	3	4	0	2	4	2	2	3	1	3	2	1	2	1											
The Canterville Ghost	3	2	4	4	4	3	1	4	2	2	1	2	3	4	1	1	1	3	3	0	3	1	2	4	1	1	3	1	3	1	3	1	2											
Cannery Row	3	3	4	4	3	0	0	1	2	1	1	0	1	2	3	1	1	2	2	0	1	1	1	3	2	1	3	0	2	1	0	1	0											
Under the Greenwood Tree	2	3	4	4	4	0	1	2	1	1	3	4	3	1	1	1	1	2	2	0	1	0	0	4	3	1	1	3	0	2	1	0	0											
The Perilous Life of Jade Yeo	1	3	3	3	3	4	1	2	2	1	1	4	4	2	2	1	1	2	2	2	1	1	4	2	1	4	3	3	2	2	0	0	0											
The Tiny Wife	3	3	1	3	4	2	3	3	3	1	1	2	2	3	4	1	1	1	4	1	4	0	2	4	4	4	4	4	4	4	4	4	3											
AVERAGE	1.03	1.76	2.39	2.91	2.74	1.54	1.83	2.80	1.80	2.00	1.60	1.74	2.69	2.91	1.23	1.46	1.83	3.31	1.66	3.09	1.89	2.23	3.83	2.43	2.14	2.94	0.97	3.03	2.29	1.40	2.44	4.3												

Figure 2: Full annotation scores for all 35 novellas across Human (H) and LLM (Claude, Gemini, DeepSeek) conditions under both direct and PatternTeller-mediated annotation. Scores are on a 0–4 scale.

D. Selection of Key Narrative Elements

Table 5

Narrative concepts ranked by structural importance (MultiTaskLasso with stability selection)

Rank	Concept	Importance (MTL)	Importance (baseline)	Count	Notes
1	Closure	0.783	0.784	1	Structural centrality and frequency
2	Final Confrontation	0.699	0.761	1	Structural centrality and frequency
3	Return	0.688	0.814	1	Structural centrality and frequency
4	Resurrection	0.662	0.517	1	Structural centrality; Hero's Journey specific
5	Denouement	0.648	0.481	1	Structural centrality
6	Confrontation of Suspects	0.610	0.716	1	Frequent; Mystery specific
7	Resolution and Aftermath	0.589	0.663	1	Frequent
8	Return with Elixir	0.588	0.357	1	Hero's Journey specific
9	Final Revelation	0.557	0.882	1	Very frequent
10	Reward	0.551	0.663	1	Frequent
11	Realization or Insight	0.547	0.906	1	Very frequent
12	Red Herrings and False Leads	0.526	0.566	1	Mystery specific
13	Meeting with Mentor	0.507	0.718	1	Romance / Hero's Journey specific
14	Initial Investigation	0.487	0.853	1	Very frequent; Mystery specific
15	Falling Action	0.472	0.707	1	Frequent
16	Development of Resistance	0.464	0.787	1	Satire specific
17	Final Act	0.459	0.498	1	Tragedy specific
18	Approach to Inmost Cave	0.454	0.787	1	Hero's Journey specific
19	Resolution	0.429	0.663	1	Frequent
20	Downfall	0.425	0.581	1	Tragedy specific
21	Uncovering Secrets and Clues	0.422	0.762	1	Mystery specific
22	Consequence of Resistance	0.419	0.726	1	Satire specific
23	Climax	0.416	0.762	1	Frequent
24	Crisis and Struggle	0.402	0.744	1	Satire specific
25	Presentation of Mystery	0.395	0.744	1	Mystery specific
26	Revelation of the System	0.380	0.765	1	Satire specific
27	Crisis	0.378	0.720	1	Frequent
28	Departure	0.357	0.689	1	Hero's Journey specific
29	Ordeal	0.344	0.689	1	Hero's Journey specific
30	Protagonist's Awakening	0.339	0.650	1	Genre-neutral
31	Call to Adventure	0.302	0.600	1	Hero's Journey specific
32	Awakening	0.224	0.451	1	Genre-neutral
33	Trials and Challenges	0.191	0.451	1	Genre-neutral
34	Establishment of Setting	0.183	0.328	1	Introductory; low centrality
35	Introduction of Characters	0.165	0.328	1	Introductory; low centrality
36	Confrontation	0.146	0.328	1	Low centrality

Notes: **Structural centrality and frequency** — high MTL importance and high baseline (top-3 concepts); **Structural centrality** — high MTL importance but low baseline (structurally defining when present, but rare); **Very frequent** — very high baseline (> 0.85) but moderate MTL importance (ubiquitous but not structurally distinctive); **Frequent** — moderately high baseline; **X specific** — concept belongs primarily to a single genre (Mystery, Satire, Tragedy, Romance, or Hero's Journey); **Introductory / low centrality** — low MTL importance, typically appearing at the narrative opening.

Table 6Structural elements ranked by stability selection (MultiTaskLasso, $n_{runs} = 200$, $sample_frac = 0.75$)

Rank	Genre	Element	Stage	Selection Rate	Mean Importance
1	Satire	Development of Resistance	4	1.000	0.948
2	Mystery	Red Herrings and False Leads	6	1.000	0.925
3	Mystery	Uncovering Secrets and Clues	4	1.000	0.920
4	Comedy	Climax	4	1.000	0.918
5	Tragedy	Crisis	5	1.000	0.911
6	Romance	Meeting with Mentor	2	1.000	0.898
7	Comedy	Falling Action	5	1.000	0.884
8	Satire	Crisis and Struggle	5	1.000	0.882
9	Mystery	Resolution and Aftermath	9	1.000	0.872
10	Romance	Approach to Inmost Cave	5	1.000	0.871
11	Tragedy	Downfall	6	1.000	0.862
12	Mystery	Final Confrontation	8	1.000	0.842
13	Mystery	Presentation of Mystery	2	1.000	0.821
14	Mystery	Initial Investigation	3	1.000	0.820
15	Romance	Return	8	1.000	0.809
16	Mystery	Confrontation of Suspects	5	1.000	0.798
17	Mystery	Realization or Insight	7	1.000	0.797
18	Romance	Reward	7	1.000	0.784
19	Comedy	Resolution	6	1.000	0.782
20	Comedy	Denouement	7	1.000	0.781
21	Satire	Consequence of Resistance	6	1.000	0.778
22	Satire	Closure	8	1.000	0.750
23	Tragedy	Final Act	7	1.000	0.725
24	Satire	Final Revelation	7	1.000	0.709
25	Romance	Resurrection	9	1.000	0.639
26	Romance	Return with Elixir	10	1.000	0.637
27	Romance	Departure	3	0.995	0.896
28	Romance	Call to Adventure	1	0.985	0.887
29	Romance	Ordeal	6	0.980	0.737
30	Satire	Protagonist's Awakening	3	0.965	0.861
31	Tragedy	Awakening	3	0.950	0.858
32	Romance	Trials and Challenges	4	0.950	0.776
33	Satire	Revelation of the System	2	0.900	0.613
34	Satire	Establishment of Setting and Characters	1	0.745	0.666
35	Tragedy	Confrontation	4	0.745	0.657
36	Mystery	Introduction of Characters and Setting	1	0.725	0.601

Notes: Horizontal rule separates elements with selection rate = 1.000 (rows 1–26) from elements with selection rate < 1.000 (rows 27–36). Selection rate is the fraction of subsampling runs ($n = 200$) in which the element was selected by MultiTaskLasso (α chosen via 5-fold MultiTaskLassoCV). Mean importance is the average ℓ_2 norm of the corresponding coefficient column across all runs.