The Emotion Dynamics of Literary Novels

Krishnapriya Vishnubhotla[1,3] and **Adam Hammond**[2] and **Graeme Hirst**[1]

[1]Department of Computer Science, University of Toronto
[2]Department of English, University of Toronto
[3]Vector Institute, Toronto
vkpriya@cs.toronto.edu, adam.hammond@mail.utoronto.ca, gh@cs.toronto.edu

Saif M. Mohammad
National Research Council Canada
saif.mohammad@nrc-cnrc.gc.ca

Abstract

Stories are rich in the emotions they exhibit in their narratives and evoke in the readers. The emotional journeys of the various characters within a story are central to their appeal. Computational analysis of the emotions of novels, however, has rarely examined the variation in the emotional trajectories of the different characters within them, instead considering the entire novel to represent a single story arc. In this work, we use character dialogue to distinguish between the emotion arcs of the narration and the various characters. We analyze the emotion arcs of the various characters in a dataset of English literary novels using the framework of Utterance Emotion Dynamics. Our findings show that the narration and the dialogue largely express disparate emotions through the course of a novel, and that the commonalities or differences in the emotional arcs of stories are more accurately captured by those associated with individual characters.

1 Introduction

Storytelling is one of the earliest of all human traditions, predating even the invention of writing. Emotions are a powerful aspect of storytelling in any form, be it the oral tradition, or with the written word in fictional novels: a compelling story is often one that is able to evoke a strong emotional reaction from its audience. Characters (fictional named entities) are often at the center of this emotional connection that readers form with a story; we identify with their struggles, react to their pain, and celebrate their victories. The different characters within a single story usually display varying emotional trajectories through the course of the narrative. This variation is also seen at the level of the novel itself: some are tragedies, some are comedies; some have very dramatic highs and lows, others have a more even emotional tone.

Kurt Vonnegut, in his now-famous lecture[1] (Vonnegut, 2009), spoke about the "shapes of stories", plotting the fluctuations of a *character's good or bad fortune* on the y-axis as a function of narrative time, from beginning to end, on the x-axis. One can assume that he was referring to the emotional journey of the protagonist of the story in his analysis. Prior work in NLP on plotting the shapes of stories, however, has considered a novel as representing a single emotional trajectory, where the dialogue of all the different characters is merged together with the narration to create one overarching novel arc (a simplification made because of the lack of annotated data in which characters are mapped to their utterances — still a challenging NLP problem). In this work, we look instead at the emotion arcs of the *individual characters* in story, distinguishing them from the arcs of the narration as well as the overall novel arc, and quantify the extent of their variation both within a single novel and across novels written by different authors.

Our goal is to quantitatively capture *longitudinal patterns* of a character's emotional states (how emotions change over time), referred to as Utterance Emotion Dynamics (UED) (Hipson and Mohammad, 2021). Through these metrics we examine the following research questions about emotion arcs and emotion change (at an aggregate level) in full-length English novels:

1. *The Emotion Dynamics of Novels*: How does the emotion change from the start of the novel to the end: overall, for just the narration, and for individual characters?

2. *Narration vs. Dialogue*: Do narration and dialogue have distinct emotion arcs? Prior work on the shapes of stories has largely glossed over the differences between characters, and considered either the entire novel to be a single trajectory, or worked with the

---

[1] https://bigthink.com/high-culture/vonnegut-shapes/

text of the narration alone. We examine the validity of this approach by quantifying the differences between the emotions expressed in the narration and the dialogue of various characters.

3. *Diversity of Character Arcs*: How diverse are the emotion arcs of the characters in a story? The variety in the emotional trajectories of characters within a novel can be quite informative about the type of story it tells. A typical hero-versus-villain story, for example, might display opposing emotional arcs for the two main characters; one that follows the adventures of a group of friends might have more closely-correlated character arcs.

4. *Effects of Gender*: Are there consistent differences between the emotion arcs of gendered character groups? How does this change with different authors?

To answer these questions we conduct experiments that make use of the following resources:

1. The *Project Dialogism Novel Corpus (PDNC)* (Vishnubhotla et al., 2022), which contains manual annotations identifying the speakers of all dialogue in 28 full-length English-language novels.

2. The NRC Valence, Arousal, and Dominance Lexicon (Mohammad, 2018a), which includes ~20,000 English words with real-valued association scores (between 0 and 1) for the three dimensions of valence (V), arousal (A), and dominance (D) dimensions.[2] A score of 1 indicates a maximum association or highest V/A/D, and a score of 0 indicates that the word is associated with the lowest V/A/D.

3. A simple, accurate, and interpretable way to generate emotion arcs from sequential text using an emotion lexicon (Teodorescu and Mohammad, 2023).

4. Metrics of Emotion Dynamics (Hollenstein, 2015; Kuppens and Verduyn, 2017), which quantify patterns of emotion change.

We find that novels, on average, express emotions high in valence and lower in arousal and dominance (0.65 vs 0.38 and 0.52). Most of the high valence and dominance is expressed in character dialogue rather than narration. We show that emotional arcs of characters are quite different from that of the narration, and from one another (average correlations close to 0); the extent of this variation also changes from novel to novel. We find that female authors write characters who express higher valence, lower arousal, and higher dominance; male characters written by male authors have the highest arousal for their utterances.

Our work sheds light on aspects of storytelling that have been under-explored in computational literary studies, and quantitatively demonstrates the importance of centering characters in order to gain a more nuanced understanding of novels. We make our data and code publicly available.[3]

2 Background and Related Work

2.1 Emotion Arcs of Novels

Much of the NLP work on literary character understanding has focused on extracting character-related information from the *narration* of a novel. Verbs and adjectives associated with the tokens that refer to character entities have been used to characterize aspects of their personality, track the narrative sequence of events associated with them, and map changing character relationships (Bamman et al., 2014; Chaturvedi et al., 2016; Iyyer et al., 2016).

Here, we instead study characters through their *utterances*. Quotations in a novel characterize an aspect of narratives called *voice* that reveals individual perspectives and points of view; this, together with *tense* and *mood*, binds together different aspects of a narrative (Piper et al., 2021). The dialogue uttered by characters to converse with one another can reveal to us aspects of their personality and their relationships that are complementary to those described in the narration. These aspects have been less studied in computational analyses of literature, primarily because identifying who said what and to whom — that is, the task of quotation attribution — is still quite a challenging NLP problem (Muzny et al., 2017; Vishnubhotla et al., 2023). Elson et al. (2010) use quoted speech to find social links between characters and model relationships between them; Sims and Bamman (2020) use attributed character speech to quantify information propagation within a narrative.

Our focus is particularly on the emotional features of character utterances, their variation through the course of the novel, and the variation of

these features across different characters and novels. Large-scale analyses of the flow of emotion-bearing words in novels as a whole has been explored since the early 2010s (Mohammad, 2012; Reagan et al., 2016; Kim et al., 2017; Ohman et al., 2024). Work on individual character dialogues has been much sparser. Nalisnick and Baird (2013) dive into emotions conveyed in character dialogue with a dataset of Shakespeare's plays; an utterance is assumed to have been uttered by the closest character mentioned. They use this to test various hypotheses on the relations between characters and their evolution, such as protagonist–antagonist pairs.

2.2 Utterance Emotion Dynamics

Emotion dynamics is a framework from psychology for measuring how an individual's emotional state changes over time (Hollenstein, 2015; Kuppens and Verduyn, 2017). It has been studied in relation to emotional, mental, and physical well-being. Hipson and Mohammad (2021) introduce a computational framework to measure emotion dynamics metrics from the emotion arc of an individual's natural language utterances over a time period. They term this framework *Utterance Emotion Dynamics (UED)*. The authors primarily characterize an emotion state as a point in the two-dimensional valence–arousal space, but state that it can be over one or more emotion dimensions. In the present work, we limit ourselves to analyzing emotions individually, in a one-dimensional space, where the temporal flow is represented on the x-axis and the emotion state values on the y-axis; this follows the work done by Vishnubhotla and Mohammad (2022), who developed UED metrics further for the one-dimensional case.[4]

Given a sequence of temporally ordered utterances, the emotion state at a time point is defined as the average emotion value of a small window of utterances (or words) uttered around that time point. This window is moved forward by one word at each step to obtain a sequence of temporally ordered emotion states (the overlapping windows lead to a smoother and more continuous arc compared to using non-overlapping adjacent windows). The *home base* for an individual is the space of emotion states, or values, that they are most likely to be found in, and is captured by the *mean* (average emotion state value), which is the center of the

home base, and the *variability* (standard deviation of the emotion states), which defines the bounds on either side. The home base can be visualized as an ellipse in the 2D space; for a single dimension, it will define a range of numerical emotion state values. Any motion of the emotion state outside of the range of this home base is termed a *displacement*. Displacements in turn are characterized by a series of metrics: their temporal length (*displacement length*), the peak emotional displacement in relation to the home base (*peak distance*), and the *rate* of rise to the peak and return to the home base. We describe the various UED metrics in more detail in Appendix A.

The emotion value of a window of words can be determined in many ways: with lexicons that associate words and phrases with a numerical score along a particular emotion dimension, or with statistical and neural models that are trained to predict an emotion score given a text span as input. Lexicon-based methods are simpler and more interpretable, and do not need domain-specific training or fine-tuning of models, making them the preferred approach in inter-disciplinary research (Öhman, 2021).

Lexicon-based methods of emotion estimation may not be highly accurate at the sentence or span-level. However, Teodorescu and Mohammad (2023) showed that these methods were able to match the true emotion *arc* of a text sequence with correlations above 0.9 when window sizes exceeded 100 words. These methods are therefore able to estimate relative changes in emotion expression with a high accuracy. Ohman et al. (2024) similarly validate the utility and accuracy of lexicon-based methods of constructing emotion arcs specifically for literary texts, while emphasizing the need to tune the lexicons for the domain of text being analyzed.

Emotion lexicons in turn have been created for many emotion dimensions, and in multiple languages (LIWC (Tausczik and Pennebaker, 2010); WordNet-Affect (Bobicev et al., 2010), SentiWord-Net (Baccianella et al., 2010), VADER (Hutto and Gilbert, 2014), and the NRC suite of emotion and affect lexicons (Mohammad and Turney, 2013; Mohammad, 2018b,a)). We therefore follow this approach in our work.

---

[4]https://github.com/Priya22/EmotionDynamics

Speaker Type		Count	#tokens
Meta	novel	28	96,973.96
	narration	28	57,995.43
Characters	major	111	6,547.91
	intermediate	113	2,025.35
	minor	585	231.99

Table 1: The number of speakers, and average number of tokens, for each group of speakers that we consider, including meta-speakers.

3 Literary Emotion Dynamics

There are several emotional trajectories of interest given a novel: one can look at the emotions of entire novel text, as has been done in prior NLP work, or consider the narration and dialogue as two text streams of interest, representing the narrator and the characters (we term these text streams as being uttered by *meta-speakers*). The dialogue can further be analyzed by considering each character's utterances as an individual text stream of interest.

We categorize the characters in a novel into three groups based on the volume of their dialogue: *major characters* are those who contribute at least 10% of the total dialogue in the novel or have at least 100 attributed quotations; *minor characters* utter fewer than 35 quotations throughout; the rest are labelled *intermediate*. Table 1 presents the number of speakers and the average number of tokens for each of these speaker groups in the PDNC.

The temporal ordering of the text in a story can in turn be viewed according to narrative time or chronological time. We order utterances by narrative time, which is of substantial interest for literary analysis as it is the order in which the reader experiences the text uttered by a character[5]. We normalize the time for each speaker to lie between the range [0,1], i.e., each speaker starts their emotion state at time point 0 and ends at time point 1, irrespective of the volume of their dialogue.[6]

We computed the emotion arc for a speaker using the parameters shown to produce high-quality emotion arcs by Teodorescu and Mohammad (2023): a rolling window size of 500 words, with the window moving forward by one word at each step until the final window subsumes the final word. In order to compute the emotion state at a particular timepoint (i.e, for a window of words), we use the NRC-VAD lexicon (Mohammad, 2018a).

---

[5]Determining chronological time in a narrative text is a separate research problem on its own.

[6]One could also conduct analyses of relative positions of when a character speaks within a novel; however, we leave that for future work. This paper focuses on patterns of change *within* the emotion arc of a character.

3.1 Validation of Emotion Arcs

While an in-domain quantitative validation of emotion arcs is ideal, we do not have an annotated dataset of emotion arcs for our dataset of character dialogue. We rely on prior work that has evaluated similar methods in different domains (tweets, movie reviews, novels), and select our hyperparameters to be comparable (Teodorescu and Mohammad, 2023; Bizzoni and Feldkamp, 2023; Ohman et al., 2024). In keeping with the recommendations of Ohman et al. (2024), and as is good practice with lexicon-based methods of sentiment analysis in general (Mohammad, 2020), we remove frequently-used terms from the NRC-VAD lexicon that can be used in a sense different from the predominant word sense that was annotated, as well as common stopwords (*have*, *will*, *may*, *said*, *saw*, etc.). Additionally, one of the authors of this work, who is also an domain expert in English literature and familiar with the novels in PDNC, qualitatively examined a subset of the constructed arcs for valence and verified that they represented the trajectory of the characters. We reproduce some of the text spans from our dataset that have the highest and lowest peak displacements in Appendix B.

3.2 Aspects of Gender

For any individual, gender is a complex identity that can vary based on their personal preferences, experiences, and social expectations. Cao and Daumé (2020) describe the various facets of gender experienced by humans and their expression in linguistic form. For the characters in PDNC, we manually annotate gender information by primarily considering the referential pronouns used for them throughout the text, along with lexical indicators such as highly-gendered names and nouns (for example, identifying *Mary Jane* as Female, and *the policeman* as Male). While gender can be expressed along a spectrum of identities, we do not find many characters in the dataset outside of the male/female binary (the exceptions are a few minor animal characters in *Alice's Adventures in Wonderland*). The same is true of the authors represented in PDNC; we rely on descriptions from Wikipedia to determine their presented gender, and do not find any who fall outside of the male/female binary (details in Appendix C). Our analysis therefore is restricted to these two categories.

Speaker	Valence		Arousal		Dominance	
	Mean	Var.	Mean	Var.	Mean	Var.
Novel (meta)	0.6472	0.0531	0.3811	0.0466	0.5192	0.0582
Narration (meta)	0.6273	0.0567	0.3857	0.0462	0.5076	0.0603
Character	0.6746	0.0295	0.3766	0.0282	0.5380	0.0355

Table 2: Aggregated Mean and variability (Var.) scores for the different types of speakers in our dataset, for each of the three emotion dimensions.
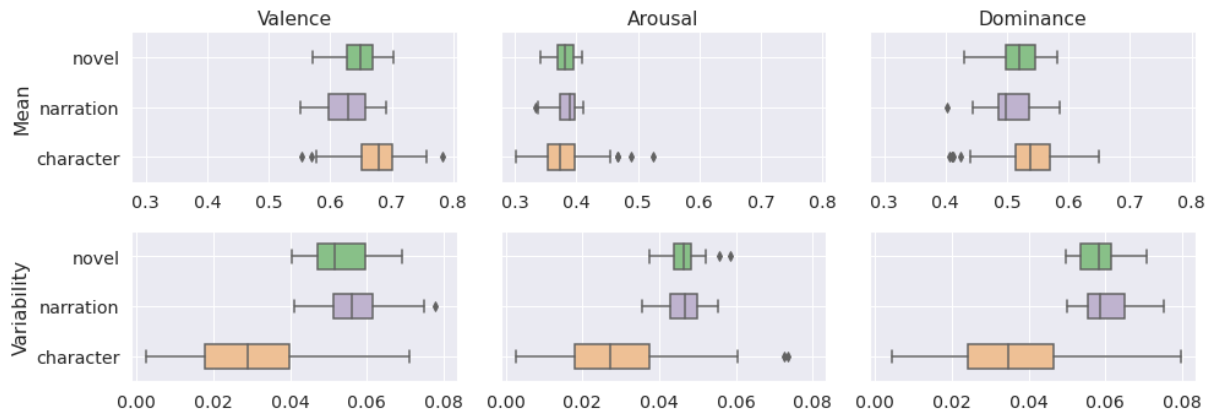

Figure 1: Boxplots of the distributions of mean and variability for each of the three affective dimensions and all novels in PDNC, when the entire novel and the narration are considered to be uttered by a single meta-speaker, and when each character's utterances are considered individually.

4 Research Questions and Experiments

We now explore a series of research questions (RQs) on the emotional dynamics of literary novels, the diversity of emotion arcs that can be found within a novel, as well as across novels, and how they relate to one another.

4.1 The Emotion Dynamics of Novels (RQ1)

We generated emotion arcs and computed UED metrics for all the novels in PDNC. Figure 1 shows the distribution of the mean and variability for the entire novel, the narration, and individual characters. Table 2 lists the aggregated scores for mean and variability. All other UED metrics are reported in Tables 5–7 in Appendix D.

It is immediately apparent that there is much more variation (i.e, the range of values covered) in the emotion dynamics of individual characters than when the narrative and/or the entire novel are clumped together into a single voice. We quantify these distinctions in detail in the next section; here, we look at the overall distributions of these metrics for our dataset.

Emotion mean: On average, novels (when considered as a single speaker) have a mean value of 0.65 for valence, 0.38 for arousal, and 0.52 for dominance. Narration alone has lower mean values of valence and dominance, and about the same for arousal. Character dialogue, on the other hand, has

a higher level of valence and dominance compared to narration — the third quartile for narration is close to the first quartile for dialogue — while the mean arousal is slightly lower, the spread of values exceeds the range for narration on either end of the distribution.

Variability: The average variability for novels (as a whole) lies around 0.053 for valence, 0.058 for dominance, and 0.046 for arousal. Narration has consistently higher average values for variability along all three dimensions when compared to character dialogue (nearly double), but the latter again covers a wider range of values.

Peak Distance and Displacement Lengths: On average, character dialogue has smaller peak distances when compared to narration across the three dimensions (0.016 vs 0.025 for valence). Displacements below the home base for valence have (significantly) higher peaks compared to those above the home base (for both narration and dialogue). Specifically for character dialogue, the average length of low valence displacements is much higher than all other displacements (106 words, compared to 87 for high valence, and an average of 87 and 82 words for arousal and dominance respectively). This indicates that the most common displacement in the emotional states of characters is when they enter more negative states.
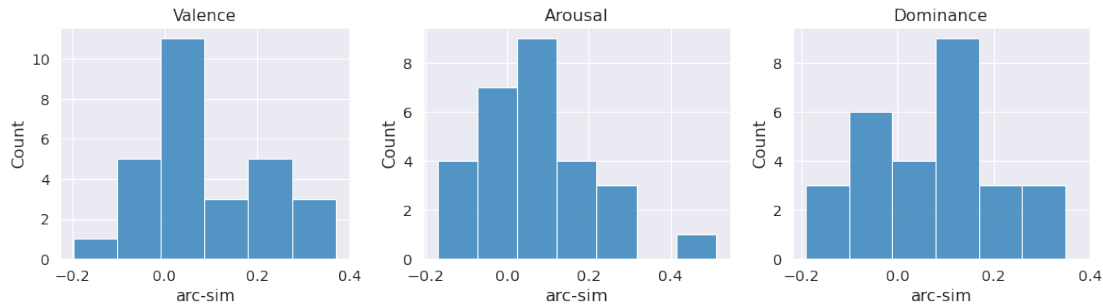
Figure 2: Distribution of arc correlations between narration and dialogue (irrespective of character) for all three VAD dimensions.

Rise and Recovery Rates: For narration, rise rates are significantly higher when compared to recovery rates, for all three dimensions, and for both low and high displacements (all metrics are reported in the Appendix Tables 5–7). For character dialogue, we do not see significant ($p < 0.05$) differences for valence (low and high displacements), and for high dominance displacements (i.e., rise and recovery rates are quite similar).

We note several outliers on either extreme for these metrics, which are potentially interesting to researchers in literary studies, as they pinpoint characters with particularly extreme and notable personality traits. For example, the character of Dr. Watson in the Sherlock Holmes novel *The Sign of the Four* has the highest rise rate for high arousal, indicating that he is easily excited (rises quickly to states of high activity/arousal). We report all meta-speakers and characters that emerge as outliers in Table 8 in Appendix E.

4.2 Emotion Arcs Within a Novel (RQ2)

In the previous section, we compared the aggregate UED metrics for the different text streams in a novel: the entire text, narration, and character dialogue. Here, we instead compare the shapes of the trajectories by considering a measure of arc similarity. We temporally align a pair of arcs that we wish to compare, and compute the Spearman correlation between them (details of the process are in Appendix F). A higher correlation score implies that the two arcs follow similar shapes.

We continue to further quantify the diversity of emotion arcs that can be found within a single story.

Q1: *How similar are the emotion arcs of the narration and dialogue in a novel (regardless of which character is uttering it)?*
This question gets at the heart of how close the emotions expressed in the narration are to those of its characters' utterances. In a way, this metric

captures a facet of narrative style — is the narrator emotionally detached from what the characters are experiencing, or does the linguistic style tend to reflect their emotional journeys?

Results: Figure 2 plots the distribution of correlation scores between arcs of narration and dialogue for all 28 novels, for each of the three emotion dimensions. On average, we see near-zero correlations across all dimensions (0.06 for arousal and dominance, 0.09 for valence), with most correlation scores falling between -0.1 and 0.35 (mild positive correlation). We note the following outliers on either side of this range: for valence, *Alice's Adventures in Wonderland* has the lowest correlation (-0.2), and *The Sport of the Gods* has the highest correlation (0.37), followed by *The Age of Innocence* (0.33). Along the dominance dimension, the arcs in *The Awakening* have a mild positive correlation (0.35); for arousal, *Daisy Miller*, a first-person narrative, is a clear outlier with a correlation of 0.51. The latter outlier is more notable in terms of the absolute value of the correlation (moderate), indicating that the emotional tone of the narration is more in tune with the dialogue.

Discussion: The correlations between the emotional states of the narration and dialogue in novels are surprisingly low; in most cases, they are near-zero. This distinction has rarely been explored in prior work on the emotional arcs of narratives, which has largely treated the entire text of the novel as presenting a singular flow of emotions (Reagan et al., 2016; Kim et al., 2017; Fudolig et al., 2022), and demonstrates that narration and character dialogue often represent distinct emotional arcs within a novel.

Q2: *How similar are the emotion arcs of the narration and each of the major characters in a novel?*
The above results tell us that the overall arcs of narration and dialogue tend to not be correlated. We now ask if, rather than being equally distant

from or close to *all* the characters, the narration tends to attach itself the emotional arc of just its protagonist(s) or antagonist(s). For each novel, we measure the similarity between the arc of the narration and each of its *major* characters.

Results: Figure 3a plots the distribution of correlation scores for major characters from all 28 novels, for the valence dimension. Arcs of narration have little to no correlation with those of the major characters (average values of 0.03, 0.02, and 0.03 for valence, arousal, and dominance respectively).[7] Figure 8a in Appendix G plots the distribution of correlations for each novel.

Interestingly, although 5 out of 28 novels in PDNC are written in the first-person[8], many of the correlations between the narration and the dialogue of the narrator fall in the mild (-0.1 to 0.2) range for valence; for arousal, *Daisy Miller* (0.37) has one of the highest correlation scores with its narrator, whereas *The Sign of The Four* (-0.36) has among the lowest. A close reading of these novels might provide critical literary insights into these results. With *The Sun Also Rises*, we have a traumatized and repressed narrator: what he says out loud is vastly different from what he thinks privately. For other novels, these results could be indicative of the difference between the emotions a character has in real time (dialogue) versus those in retrospect (narration).

Discussion: These findings reinforce the distinctions between the narration arc and those of a novel's main characters — we average at near-zero correlations for all three emotion dimensions, highlighting that any computational modeling of the emotion arcs of stories should consider the distinctions between these facets of a narrative.

4.3 Diversity of Character Arcs (RQ3)

We now focus solely on the emotion arcs of character dialogue, and quantify their variation in literary novels. We look at this variation both within a single novel and when compared across stories. These measures inform us of the diversity of emotional trajectories that a character can follow, and gets closer to the question of the "basic shapes of stories" (i.e, of character journeys) that many prior works in NLP have attempted to quantify.

Q3: *What is the average similarity of character arcs in a novel? Where do we see outliers?*
For each pair of major characters within a novel, we compute the similarity of their emotion arcs. Apart from looking at individual pairs of characters, we compute the mean and variance of the scores for all character pairs within a novel — which ones have the most diverse emotional journeys for their characters and which the least?
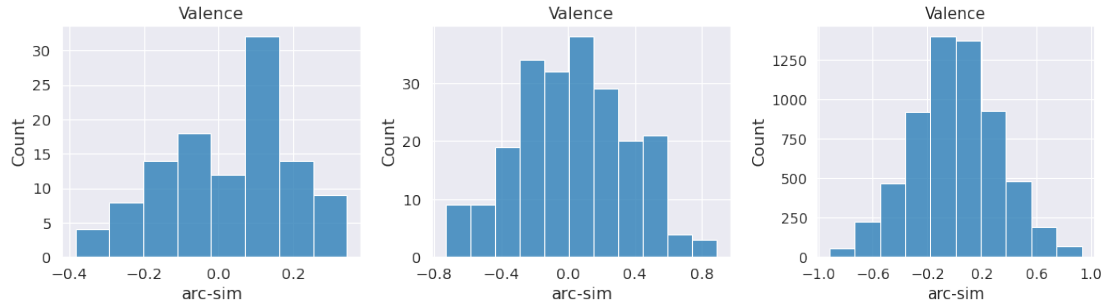
Results: We plot the distribution of all pairwise scores for valence in Figure 3b. We can immediately see the much larger spread of the scores here (than for narration and dialogue), extending into the range of high correlation and anti-correlation (maximum of 0.89 and minimum of -0.73). The mean correlations, however, stay around 0 (0.03, 0.02, 0.04 for the VAD dimensions).

We quantify the diversity of character arcs within a novel as the standard deviation of the emotion arc correlations of its major characters. Figure 8b in Appendix G shows the distribution of correlations for valence by novel. *Winnie-the-Pooh* and *The Sport of the Gods* have some of the highest diversity in major character arcs for all three dimensions; *The Age of Innocence* has the lowest diversity in valence arcs, and *The Gambler* and *The Invisible Man* for arousal. *The Sun Also Rises* leans the most positive with a median score of 0.44. In this novel, the characters Robert Cohn and Jake Barnes are often read as rivals (they are competing for the love of the same woman), but our analysis shows them travelling the same emotional journey (correlation of 0.89 for valence).

Discussion: The above results lend credence to the variety and diversity that can be found in novels at the level of characters. It is quite rare to find a story where the character trajectories or voices are uniform. While it might seem trivial to state that a story is not so much a singular narrative as a collection of intersecting narratives, computational analysis has largely suffered by not being able to afford this complexity in its study of stories. The higher correlations between character arcs as opposed to with the narration indicate that using a single novel-based arc is a poor representation of the novel for much literary analysis.

Q4: *How similar are the emotion arcs of characters across novels?*
This question explores whether we see high correlations reflecting the "shapes of stories" as described by Kurt Vonnegut, which largely describes the emo-

---

[7] Since correlation is a bipolar scale, we also compute average values for positive and negative correlations separately; none of them go beyond 0.18.

[8] *Daisy Miller*, *The Mysterious Affair at Styles*, *The Sun Also Rises*, *The Sign of the Four*, and *The Gambler*.

(a) Between narration and a major character (from the same novel).

(b) Between major characters within a novel.

(c) Between major characters across all novels.

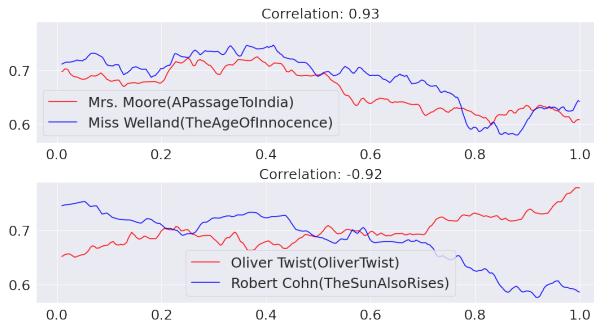Figure 3: Distribution of valence arc correlations.

Figure 4: Emotion arcs of valence for character pairs with the highest and lowest correlation scores.

tional journey of the protagonists of stories. We compute correlation scores between the arcs of all possible pairs of the 111 major characters from all 28 novels (6105 pairs).

Results: Figure 3c plots the distribution of correlation scores for valence (we see similar distributions for arousal and dominance). The average correlation between any pair of characters is close to 0, for all three dimensions. The range of correlation scores is much larger, from a minimum of -0.92 (Oliver Twist from *Oliver Twist* and Robert Cohn from *The Sun Also Rises*) to a maximum of 0.93 (Mrs. Moore from *A Passage To India* and Miss Welland from *The Age of Innocence*). In Figure 4, we plot the arcs of these two pairs of characters.

Discussion: The range of the correlations, extending from highly correlated to anti-correlated, demonstrate that character trajectories in different novels can indeed follow similar shapes, though the extent of this similarity varies widely.

4.4 Character Groups (RQ4)

Several prior works on stories have demonstrated biases in the portrayals of characters that correspond to overgeneralized and often inaccurate stereotypes of certain demographic groups, particularly those that are marginalized (Fast et al.,

2016; Sap et al., 2017). We investigate the presence of such biases for the characters in our dataset along the gender dimension, quantifying differences between the UED metrics of Male (M) and Female (F) character groups and how they are written by Male and Female authors. We note that the rather restricted range of novels and authors represented in PDNC[9] means we cannot make generalizable claims about these differences; however, our methodology is broadly applicable to any corpus of novels, and provides metrics that are useful to quantify biases in such corpora.

Q5: *Do the emotional dynamics of characters differ based on their (presented) gender?*
We have 89 Female and 121 Male characters in PDNC with sufficient utterances to compute UED metrics. We test for statistically significant differences in the aggregate UED metrics (mean, variability, average rise and recovery rates, etc.) for each of these groups using a two-sided independent t-test, and apply the Benjamini-Hochberg correction for multiple comparisons.

Results: We find the following significant ($p < 0.05$) effects: mean valence is higher for female characters compared to male characters (0.68 vs 0.66); mean arousal is higher for male characters (0.39 vs 0.37), and the average peak distance for arousal displacements is also higher for male characters compared to female characters (0.019 vs 0.016).

Discussion: This is in line with prior work on gender biases in storytelling — whether in novels, movies, or other forms of stories — that finds that female characters tend to be portrayed with higher levels of positive emotions: warm, kind, caring, and more joyful (Ramakrishna et al., 2017; Xu

---

[9]These are canonically well-regarded authors and novels, who might not be representative of the general fiction landscape of the era, and certainly not of other time-periods.
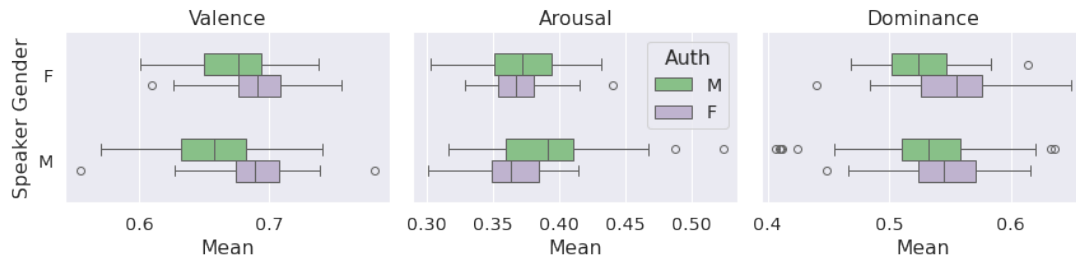
Figure 5: Distribution of the emotion mean for VAD, grouped by speaker (y-axis) and author gender (box color).

et al., 2019), while male characters are expected to express more intense emotions relating to arousal — anger, rage, and violence.

Q6: *Do male authors write their characters differently from female authors?*

We now compare the UED metrics of characters when grouped by author gender. There are 134 characters written by female authors (85 female, 49 male), and 183 characters written by male authors (133 male, 50 female). We test for significant differences in aggregate UED metrics for all possible groups of (M/F) characters written by (M/F) authors using the two-way ANOVA test.

Results: Figure 5 plots the distributions of mean VAD scores for all non-minor characters in our dataset, separated by speaker and author gender. Character dialogue written by female authors has a higher mean valence (0.69 vs 0.65), lower mean arousal (0.36 vs 0.38), and higher mean dominance (0.55 vs 0.52) when compared to that written by male authors in our dataset ($p < 0.001$). Additionally, we find that the peak distance for low arousal displacements is lower for female-authored characters (0.014 vs 0.019). Male-authored male characters have a particularly high mean arousal when compared to all other groups.

Discussion: These findings are again in line with prior work on gendered character dialogue in movie scripts and stories (Fast et al., 2016; Xu et al., 2019; Lettieri et al., 2023), which finds that female writers tend to use words that are more positively valenced, and male writers, words that are marked as more arousing. We find significant trends in UED metrics that capture the intensity of emotional displacements, which have not previously been studied. The depiction of the two character gender groups for each of the author groups also reveals trends of interest — male characters have a higher average arousal in novels authored by male writers, whereas female writers tend to send their male characters into more intense low arousal states.

5 Conclusion

In this work, we conducted a detailed analysis of the nuances of the emotional trajectories of stories. We first demonstrated that the emotions expressed in the narration of a story are not representative of those expressed by its characters. We then showed that the characters *within* and *across* stories can have widely varying emotion arcs, with correlations ranging from highly negative to strongly correlated, and averaging at near-zero correlations. These distinctions highlight the diversity of narrative threads contained in a single novel: different characters go through wildly varying journeys, and these cannot be summarized by a single emotion arc — of the novel, narration, or dialogue — alone, or even with a small inventory of prototypical shapes.

Limitations

The works that we studied here are set apart by an obvious selection bias, in that the PDNC dataset consists of only 28 novels, chosen because they are popular and critically-acclaimed in the literary canon, written by authors who were societally positioned to be able to achieve fame and success. Will we find more evident, consistent patterns of bias conditioned on character and author gender if we expand our selection to a less-curated, and therefore more representative, dataset of texts? Perhaps, but automatically extracting information about the number, gender, and types of characters in novels is computationally quite a challenging problem, as is the task of identifying and attributing the various lines of quotation within a novel to one of these characters. We hope progress in these areas of NLP research will enable more expansive analyses of literary characters in the future.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

David Bamman, Ted Underwood, and Noah A Smith. 2014. A Bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379.

Yuri Bizzoni and Pascale Feldkamp. 2023. Comparing transformer and dictionary-based sentiment models for literary texts: Hemingway as a case-study. In *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*, pages 219–228, Tokyo, Japan. Association for Computational Linguistics.

Victoria Bobicev, Victoria Maxim, Tatiana Prodan, Natalia Burciu, and Victoria Angheluş. 2010. Emotions in words: Developing a multilingual WordNet-Affect. In *Conference on Intelligent Text Processing and Computational Linguistics*.

Yang Trista Cao and Hal Daumé, III. 2020. Toward gender-inclusive coreference resolution. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4568–4595.

Snigdha Chaturvedi, Shashank Srivastava, Hal Daumé, and Chris Dyer. 2016. Modeling evolving relationships between characters in literary novels. In *AAAI Conference on Artificial Intelligence*.

David K. Elson, Nicholas Dames, and Kathleen McKeown. 2010. Extracting social networks from literary fiction. In *Annual Meeting of the Association for Computational Linguistics*.

Ethan Fast, Tina Vachovsky, and Michael S. Bernstein. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. *ArXiv*, abs/1603.08832.

Mikaela Irene D. Fudolig, T. Alshaabi, Kathryn Cramer, Christopher M. Danforth, and Peter Sheridan Dodds. 2022. A decomposition of book structure through ousiometric fluctuations in cumulative word-time. *Humanities and Social Sciences Communications*, 10:1–12.

Will E. Hipson and Saif M. Mohammad. 2021. Emotion dynamics in movie dialogues. *PLoS ONE*, 16:1–19.

Tom Hollenstein. 2015. This time, it's real: Affective flexibility, time scales, feedback loops, and the regulation of emotion. *Emotion Review*, 7:308 – 315.

Clayton J. Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan L. Boyd-Graber, and Hal Daumé. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *North American Chapter of the Association for Computational Linguistics*.

Evgeny Kim, Sebastian Padó, and Roman Klinger. 2017. Investigating the relationship between literary genres and emotional plot development. In *LaTeCH@ACL*.

Peter Kuppens and Philippe Verduyn. 2017. Emotion dynamics. *Current Opinion in Psychology*, 17:22–26. Emotion.

Giada Lettieri, Giacomo Handjaras, Erika Bucci, Pietro Pietrini, and Luca Cecchetti. 2023. How male and female literary authors write about affect across cultures and over historical periods. *Affective Science*.

Saif M. Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53:730–741.

Saif M. Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Saif M. Mohammad. 2018b. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.

Saif M. Mohammad. 2020. Practical and ethical considerations in the effective use of emotion and sentiment lexicons. *arXiv:2011.03492*.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Grace Muzny, Michael Fang, Angel Chang, and Dan Jurafsky. 2017. A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 460–470.

Eric T. Nalisnick and Henry S. Baird. 2013. Character-to-character sentiment analysis in Shakespeare's plays. In *Annual Meeting of the Association for Computational Linguistics*.

Emily Öhman. 2021. The validity of lexicon-based sentiment analysis in interdisciplinary research. In *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*, pages 7–12.

Emily Ohman, Yuri Bizzoni, Pascale Feldkamp Moreira, and Kristoffer Nielbo. 2024. EmotionArcs: Emotion arcs for 9,000 literary texts. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 51–66, St. Julians, Malta. Association for Computational Linguistics.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311.

Anil Ramakrishna, Victor R. Martinez, Nikos Malandrakis, Karan Singla, and Shrikanth S. Narayanan. 2017. Linguistic analysis of differences in portrayal of movie characters. In *Annual Meeting of the Association for Computational Linguistics*.

Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5.

Maarten Sap, Marcella Cindy Prasetio, Ari Holtzman, Hannah Rashkin, and Yejin Choi. 2017. Connotation frames of power and agency in modern films. In *EMNLP*.

Matthew Sims and David Bamman. 2020. Measuring information propagation in literary social networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 642–652.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29:24 – 54.

Daniel Teodorescu and Saif M. Mohammad. 2023. Evaluating emotion arcs across languages: Bridging the global divide in sentiment analysis.

Krishnapriya Vishnubhotla, Adam Hammond, and Graeme Hirst. 2022. The Project Dialogism Novel Corpus: A dataset for quotation attribution in literary texts. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5838–5848, Marseille, France. European Language Resources Association.

Krishnapriya Vishnubhotla and Saif M. Mohammad. 2022. Tweet emotion dynamics: Emotion word usage in tweets from US and Canada. In *Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France.

Krishnapriya Vishnubhotla, Frank Rudzicz, Graeme Hirst, and Adam Hammond. 2023. Improving automatic quotation attribution in literary novels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 737–746.

Kurt Vonnegut. 2009. *Palm Sunday: An Autobiographical Collage*. Random House Publishing Group.

Huimin Xu, Zhang Zhang, Lingfei Wu, and Cheng-Jun Wang. 2019. The Cinderella complex: Word embeddings reveal gender stereotypes in movies and books. *PloS ONE*, 14(11):e0225385.

A Utterance Emotion Dynamics

The Utterance Emotion Dynamics framework derives several metrics characterizing the temporal patterns of regularity and change of emotion states derived from the textual utterances of an individual. Figure 6 shows a simple example emotion arc and the UED metrics corresponding to the home base and displacements below and above the home base. We briefly describe these metrics here:

- **Emotion mean (emo_mean)**: The mean of the sequence of emotion states.
- **Variability (emo_std)**: The standard deviation of the sequence of emotion states.
- **Average peak distance (emo_avg_peak_dist)**: Average of the peak emotional distance from the home base for all displacements (a measure of how emotional the speaker gets on average when they have a displacement).
- **Average displacement length (emo_avg_disp_length)**: Average of the length of a displacement, in terms of temporal steps, for all displacements (a measure of how long the speaker is outside the home base per displacement on average).
- **Average rise rate (emo_rise_rate)**: Average of the rise rates of all displacements (a measure of how quickly one reaches peak distance from home base, regardless of direction of displacement).
- **Average recovery rate (emo_recovery_rate)**: Average of the recovery rates of all displacements (a measure of how quickly one reaches peak distance from home base (regardless of direction of displacement)).
- **Average Low peak distance (emo_low_peak_dist)**: Average of the peak emotional distance from the home base for all displacements below the home base.
- **Average Low displacement length (emo_low_disp_length)**: Average of the length of a displacement, in terms of temporal steps, for all displacements below the home base.
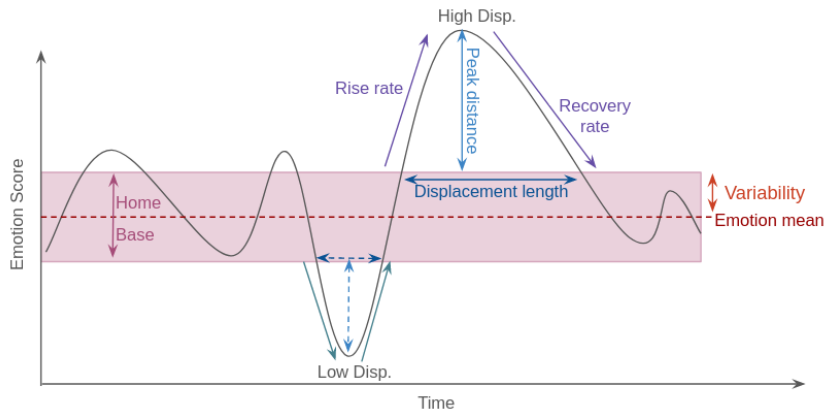
Figure 6: A visual representation of the various UED metrics for a sample emotion arc.

- **Average Home-to-Low rise rate (emo_low_rise_rate)**: Average of the rise rates of all displacements below the home base (a measure of how quickly one descends to the lowest emotion state).
- **Average Low-to-Home recovery rate (emo_low_recovery_rate)**: Average of the recovery rates of all displacements below the home base (a measure of how quickly one recovers from the lowest emotion state).
- **Average High peak distance (emo_high_peak_dist)**: Average of the peak emotional distance from the home base for all displacements above the home base.
- **Average High displacement length (emo_high_disp_length)**: Average of the length of a displacement, in terms of temporal steps, for all displacements above the home base.
- **Average Home-to-High rise rate (emo_high_rise_rate)**: Average of the rise rates of all displacements above the home base (a measure of how quickly one ascends to the lowest emotion state).
- **Average High-to-Home recovery rate (emo_high_recovery_rate)**: Average of the recovery rates of all displacements above the home base (a measure of how quickly one recovers from the highest emotion state).

B Words Behind the Arcs

In Table 3, we reproduce some of the spans of character dialogue from the PDNC corpus with the highest and lowest displacements in their emotion arcs along the valence dimension. The lowest point for Jake Barnes from the *The Sun Also Rises*, for example, occurs when the character finds himself drunk, alone, and in a fight with his friend, Robert

Cohn (at around the 80% mark in the novel).

C Dataset: PDNC

We list details of the novels in the Project Dialogism Novel Corpus in Table 4. All details were annotated by the dataset creators (Vishnubhotla et al., 2022) and were made publicly available at https://github.com/Priya22/project-dialogism-novel-corpus.

D UED Metrics

We report the aggregated UED metrics for the different types of speakers -=- novel (meta-speaker), narration (meta-speaker), all characters, major characters, intermediate characters, and minor characters -=- for valence (Table 5), arousal (Table 6), and dominance (Table 7).

E Outlier Characters

Speakers who emerge as outliers for the emotion mean and variability metrics, along all three VAD dimensions, are listed in Table 8. Speakers can include meta-speakers (such as narration). Outliers are identified as points that fall outside the whiskers of a box-and-whisker plot, i.e, the scores that are below $Q_1 - 1.5 \times IQR$ (low) and above $Q_3 + 1.5 \times IQR$ (high), where Q_1 and Q_3 are the 25th and 75th percentiles (1st and 3rd quartiles) of the distribution, and $IQR = Q_3 - Q_1$ is the inter-quartile range. Outliers are independently identified for each speaker type (novel, narration, character).

F Aligning Emotion Arcs

Let's say we want to compare the computed emotion arcs of a set of speakers S:

Speaker	Novel	Emo State	Utterance Span
Jake Barnes	The Sun Also Rises	Low	"Oh, to hell with Cohn, .. Up to Paris. ... Oh, to hell with him. ... Pick up with his old girl, probably. ... Sure, I'd believe anything. Including nightmares. ... Low as hell. I feel like hell, I feel like hell. It won't do any good. Fine. .. I feel tight. That's what you wanted, wasn't it? Well, I'm tight. Is that what you want? I won't sit down, I'm going over to the hotel. ... I'm drunk, I'm going in and lie down. Yes, I'm blind. ... I got hungry and woke up."
The Invisible Man	The Invisible Man	Low	"It was night outside long before the business was over, and nothing was to be seen but the dim eyes and the claws. I stopped the gas engine, felt for and stroked the beast, which was still insensible, and then, being tired, left it sleeping on the invisible pillow and went to bed. I found it hard to sleep. I lay awake thinking weak aimless stuff, going over the experiment over and over again, or dreaming feverishly of things growing misty and vanishing about me, until everything, the ground I stood on, vanished, and so I came to that sickly falling nightmare one gets. ... I never saw any more of it."
Miss Bates	Emma	High	"So obliging of you! No, we should not have heard, if it had not been for this particular circumstance, of her being to come here so soon. My mother is so delighted! for she is to be three months with us at least. Three months, she says so, positively, as I am going to have the pleasure of reading to you. ... But ever since we had the history of that day, I have been so fond of Mr. Dixon!"
Winnie-the-pooh	Winnie The Pooh	High	"I'm very glad, that I thought of giving you a Useful Pot to put things in. ... I know, I suppose it's just practice, I wonder if Kanga will have to practise too so as to understand it. ... You mean like telling her a little bit of poetry or something? Good afternoon, Kanga. Hallo, Roo, my little fellow! Oh, Kanga, I don't know if you are interested in Poetry at all? Oh! Talking of Poetry, I made up a little piece as I was coming along. It went like this. Er now let me see"

Table 3: Sample text spans of dialogue from different characters in the PDNC corpus that were quantified as extreme emotional displacements, either into the emotion states lower than the home base (Low), or higher than the home base (High). Note that these text spans are not necessarily contiguous dialogue blocks from the novel text, but may be interrupted by the dialogue of other characters and/or the narration (indicated by the ellipsis); however, we consolidate all the utterances of a character into a single temporally-ordered text stream in order to construct emotion arcs.

Novel Title	Narrator Gender	Genre	Author Name	Author Gender
A Handful Of Dust	U	literary	Evelyn Waugh	M
Alice's Adventures in Wonderland	U	children	Lewis Carroll	M
Anne Of Green Gables	U	children	Lucy Maud Montgomery	F
A Passage to India	U	literary	Edward Morgan Forster	M
A Room With A View	U	literary	Edward Morgan Forster	M
Daisy Miller	M	literary	Henry James	M
Emma	A	literary	Jane Austen	F
Hard Times	U	literary	Charles Dickens	M
Howards End	M	literary	Edward Morgan Forster	M
Mansfield Park	A	literary	Jane Austen	F
Night and Day	U	literary	Virginia Woolf	F
Northanger Abbey	A	literary	Jane Austen	F
Oliver Twist	U	literary	Charles Dickens	M
Persuasion	A	literary	Jane Austen	F
Pride and Prejudice	A	literary	Jane Austen	F
Sense and Sensibility	A	literary	Jane Austen	F
The Age of Innocence	U	literary	Edith Wharton	F
The Awakening	U	literary	Kate Chopin	F
The Gambler	M	literary	Fyodor Mikhailovich Dostoevsky	M
The Invisible Man	U	scifi	Herbert George Wells	M
The Man Who Was Thursday	U	literary	Gilbert Keith Chesterton	M
The Mysterious Affair At Styles	M	crime	Agatha Christie	F
The Picture Of Dorian Gray	U	literary	Oscar Wilde	M
The Sign of the Four	M	crime	Arthur Conan Doyle	M
The Sport of the Gods	U	literary	Paul Laurence Dunbar	M
The Sun Also Rises	M	literary	Ernest Hemingway	M
Where Angels Fear to Tread	U	literary	Edward Morgan Forster	M
Winnie-The-Pooh	M	children	Alan Alexander Milne	M

Table 4: Metadata for the 28 novels in PDNC, which we use in our experiments.

metric // speaker_type	novel	narration	character	major	intermediate	minor
emo_mean	0.647	0.627	0.675	0.667	0.678	0.684
emo_std	0.053	0.057	0.029	0.038	0.026	0.018
emo_avg_peak_dist	0.023	0.025	0.016	0.019	0.015	0.012
emo_avg_disp_length	132.422	134.792	78.854	101.728	70.969	44.377
emo_rise_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_low_peak_dist	0.027	0.028	0.021	0.025	0.020	0.012
emo_low_disp_length	137.656	140.427	105.963	135.199	100.196	48.943
emo_low_rise_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_low_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_high_peak_dist	0.020	0.022	0.016	0.018	0.015	0.014
emo_high_disp_length	130.319	136.301	87.169	110.648	76.490	55.121
emo_high_rise_rate	0.001	0.001	0.001	0.001	0.001	0.002
emo_high_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001

Table 5: Averaged UED metrics (rows) of **valence** for the different speaker types (columns) in PDNC.

metric // speaker_type	novel	narration	character	major	intermediate	minor
emo_mean	0.381	0.384	0.377	0.381	0.377	0.366
emo_std	0.047	0.046	0.028	0.037	0.024	0.016
emo_avg_peak_dist	0.020	0.021	0.016	0.021	0.014	0.011
emo_avg_disp_length	120.911	121.039	78.358	106.449	66.519	39.966
emo_rise_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_low_peak_dist	0.018	0.018	0.017	0.020	0.015	0.012
emo_low_disp_length	126.509	120.441	89.166	118.320	75.781	49.623
emo_low_rise_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_low_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_high_peak_dist	0.023	0.025	0.019	0.024	0.016	0.011
emo_high_disp_length	118.200	125.468	85.401	117.221	71.227	40.459
emo_high_rise_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_high_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001

Table 6: Averaged UED metrics (rows) of **arousal** for the different speaker types (columns) in PDNC.

metric // speaker_type	novel	narration	character	major	intermediate	minor
emo_mean	0.519	0.508	0.538	0.532	0.545	0.536
emo_std	0.058	0.060	0.036	0.045	0.032	0.022
emo_avg_peak_dist	0.025	0.026	0.020	0.023	0.020	0.015
emo_avg_disp_length	121.263	124.919	70.747	91.681	63.142	38.689
emo_rise_rate	0.001	0.001	0.002	0.002	0.002	0.002
emo_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.002
emo_low_peak_dist	0.026	0.025	0.023	0.027	0.022	0.014
emo_low_disp_length	124.514	126.597	86.892	116.789	74.472	39.135
emo_low_rise_rate	0.001	0.001	0.002	0.001	0.002	0.002
emo_low_recovery_rate	0.001	0.001	0.001	0.001	0.001	0.001
emo_high_peak_dist	0.025	0.026	0.021	0.023	0.021	0.018
emo_high_disp_length	121.133	126.999	77.927	94.625	71.660	51.979
emo_high_rise_rate	0.001	0.001	0.002	0.002	0.002	0.002
emo_high_recovery_rate	0.001	0.001	0.001	0.001	0.002	0.002

Table 7: Averaged UED metrics (rows) of **dominance** for the different speaker types (columns) in PDNC.

dim	metric	meta	extreme	name	value
valence	emo_std	narration	high	narrator (ThePictureOfDorianGray)	0.078
valence	emo_mean	character	low	Monks (OliverTwist)	0.580
valence	emo_mean	character	low	Mr. Bumble (OliverTwist)	0.599
valence	emo_mean	character	low	The Invisible Man (TheInvisibleMan)	0.578
valence	emo_mean	character	low	Mike Campbell (TheSunAlsoRises)	0.578
arousal	emo_mean	narration	low	narrator (WinnieThePooh)	0.334
arousal	emo_mean	character	high	Professor De Worms (TheManWhoWasThursday)	0.452
arousal	emo_mean	character	high	Joe Hamilton (TheSportOfTheGods)	0.488
arousal	emo_std	character	low	Jock Grant-Menzies (AHandfulOfDust)	0.008
arousal	emo_std	character	low	Cassandra Otway (NightAndDay)	0.013
arousal	emo_std	character	high	Prince Charming (ThePictureOfDorianGray)	0.061
arousal	emo_std	character	high	Mike Campbell (TheSunAlsoRises)	0.074
arousal	emo_std	character	high	Piglet (WinnieThePooh)	0.072
dominance	emo_mean	narration	low	narrator (WinnieThePooh)	0.403
dominance	emo_mean	character	low	John Andrew (AHandfulOfDust)	0.412
dominance	emo_mean	character	low	Eeyore (WinnieThePooh)	0.407
dominance	emo_std	character	low	Joe Hamilton (TheSportOfTheGods)	0.008
dominance	emo_std	character	high	Mary Musgrove (Persuasion)	0.075
dominance	emo_std	character	high	Christopher Robin - Story (WinnieThePooh)	0.080

Table 8: Outliers among characters and narration on either extreme for the mean and variability metrics, along all three dimensions.

- Find the speaker $s_{sm} \in S$ with the smallest temporal length (i.e, the fewest utterances).

- Start with an initial timestep window of [0, 0.001] for this speaker (this initial window size is a hyperparameter we examine later).

- Move forward this window by one word for s_{sm}. This corresponds to a new bin of timestep values for the speaker s_{sm}, say [0.001, 0.011].

- Continue to move forward the window by one word until the end of the speaker's arc; this results in a set of n time bins.

- Average the emotion state values at the timesteps contained within each the n time bins to obtain n comparable emotion values for each speaker in S.

Note that, while the window is moving forward by 1 word for s_{sm}, it could be moving forward by k words for a different speaker s_j who has a longer volume of utterances. However, this ensures that we have an equal number of bins for each speaker, at approximately the same relative time points (i.e, the 1% mark, the 1.1% mark, etc).

Once a pair of emotion arcs has been aligned, we obtain equal-sized sequences of emotion states for the normalized bins, for all the speaker arcs we wish to compare (using a similarity metric such as the euclidean distance or Spearman correlation).

Qualitative Assessment: How well do the aligned emotion arcs represent the original UED arc of a novel? We qualitatively examine the aligned arcs with three different initial window sizes. The smallest novel out of the set of 28 in PDNC is *Daisy Miller*. With an initial window size of [0, 0.01], we get 22,603 temporal bins; for an initial window of [0, 0.001], we have 22,809 bins; [0, 0.05] yields 21,690 bins. We therefore compute a new set of aligned arcs for each novel by averaging the emotion state values within each of these temporal bins, for all three temporal bin sizes.

In figure 7, we plot the original UED arc and the aligned arc for a longer novel, *A Room With A View*, for each of these choices. While a bin size of [0, 0.001] retains many of the sharp transitions of the original arc, and [0, 0.05] blurs over too many of them, [0, 0.01] seems to be a good compromise in roughly capturing the ups and downs of the emotion arc. We do note, however, that the appropriate bin size is dictated by the research question of interest; if we wanted to obtain a high-level view of the "shape of a story", we might prefer to smooth the arc even further.
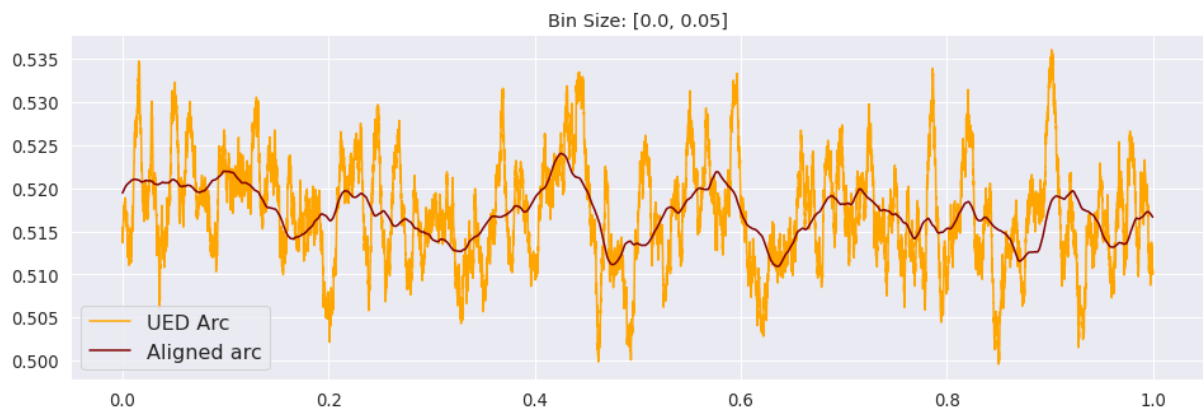
G Arc Correlations by Novel

In Figure 8a, we plot for each novel the distributions of the correlations between the arcs of each major character and the narration. In Figure 8b, we report the distributions of the correlations between the arcs of pairs of major characters within each novel.
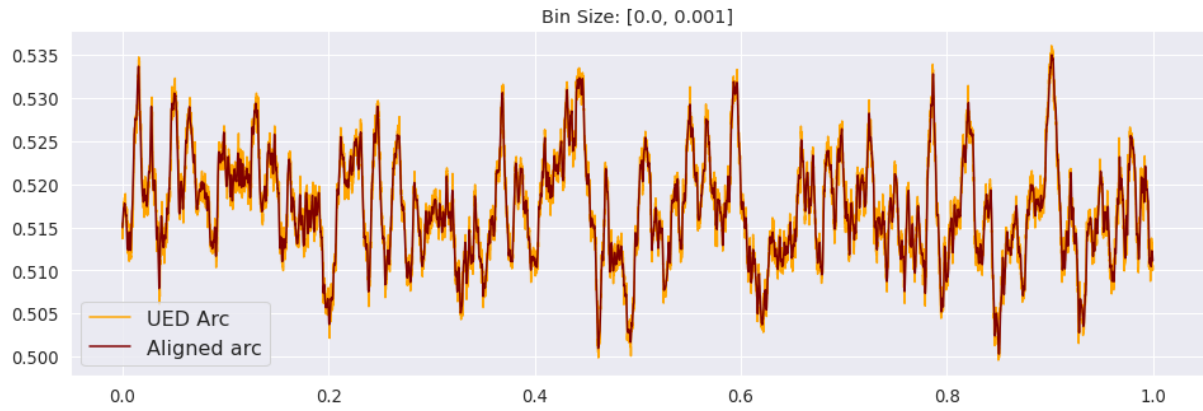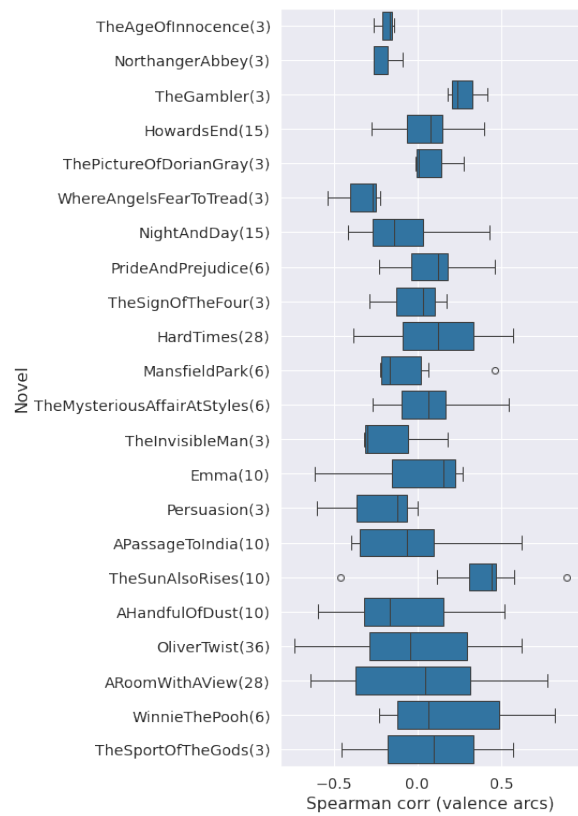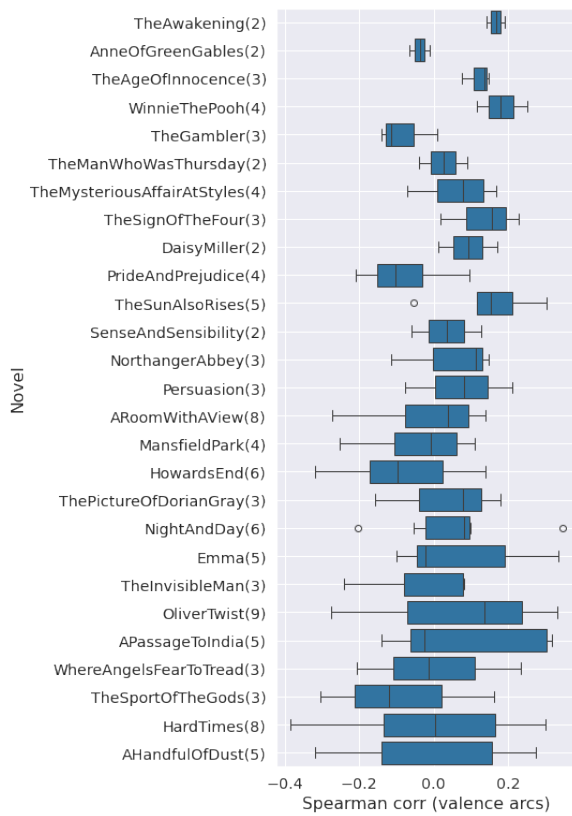
(a) Initial bin size [0, 0.001]
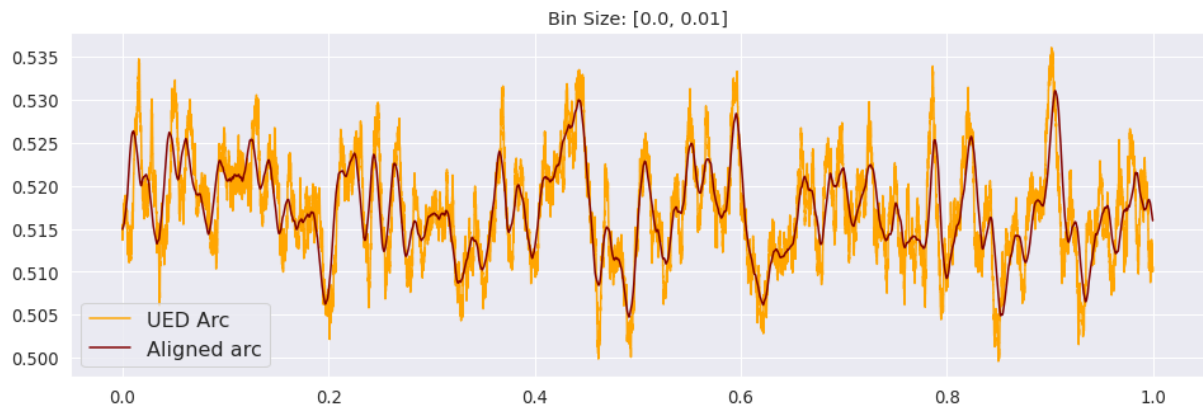
(b) Initial bin size [0, 0.01]

(c) Initial bin size [0, 0.05]

Figure 7: A visual comparison of the original UED arc with the time-aligned arc for *A Room With A View* (when aligned with the smallest novel in our dataset, *Daisy Miller*) with different initial bin sizes.

(a) Boxplots of the correlation scores of narration-major character valence arcs for each novel (ordered by variance).

(b) Boxplots of the major character valence arc correlations by novel (ordered by variance).

Figure 8: Within-novel correlations of **valence** arcs between narration and major character arcs. The numbers in parentheses beside each novel indicate the number of pairwise correlations that are represented (based on the number of major characters) for each novel.