

Article

# From Review to Genre to Novel and Back

## An Attempt to Relate Reader Impact to Phenomena of Novel Text

Marijn Koolen<sup>1</sup>   
Joris J. van Zundert<sup>2</sup>   
Eva Viviani<sup>3</sup>   
Carsten Schnober<sup>3</sup>   
Willem van Hage<sup>3</sup>   
Katja Tereshko<sup>2</sup> 

1. DHLab, Humanities Cluster, Amsterdam, The Netherlands.
2. Computational Literary Research, Huygens Institute , Amsterdam, The Netherlands.
3. eScience Center , Amsterdam, The Netherlands.

**Citation**

Marijn Koolen, Joris van Zundert, Eva Viviani, Carsten Schnober, Willem van Hage, and Katja Tereshko (2024). "From Review to Genre to Novel and Back. An Attempt to Relate Reader Impact to Phenomena of Novel Text". In: *Journal of Computational Literary Studies* 3 (1). [10.48694/jcls.3927](https://doi.org/10.48694/jcls.3927)

**Date published** 2024-10-17

**Date accepted** 2024-09-28

**Date received** 2024-01-25

**Keywords**

reading impact, literary novels, genre, topic modeling

**License**

CC BY 4.0 

**Reviewers**

Keli Du, Roxana Patras

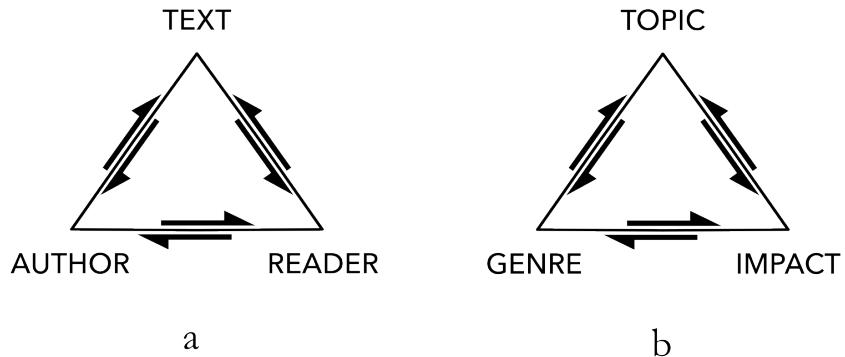
**Note**

This paper has passed through the conference track of JCLS. In addition to being peer reviewed, it was presented and discussed at the 3rd Annual Conference of Computational Literary Studies at Vienna, Austria, in June 2024.

**Abstract.** We are interested in the textual features that correlate with the reported impact by readers of novels. We operationalize impact measurement through a rule-based reading impact model and apply it to 634,614 reader reviews mined from seven review platforms. We compute co-occurrences of impact-related terms and their keyness for genres represented in the corpus. The corpus consists of the full text of 18,885 books from which we derived topic models. The topics we find correlate strongly with genre, and we get strong indicators for which key impact terms are connected to which genre. These key impact terms give us a first evidence-based insight into genre-related readers' motivations.

## 1. Introduction

Already Aristotle noted the reciprocal relations between an author, the text the author creates, and the response from an audience to the text. This fundamental model of rhetorical poetics has remained relevant throughout the ages (see e.g., Abrams 1971; Warnock 1978). The dynamics of the relations between author, text, and reader have been heavily theorized and fiercely debated (see e.g., Hickman 2012; Wimsatt 1954). But if there is no lack of theory, it appears to be much harder to gain empirical insights into these relations, though not for lack of trying by practitioners in such fields as empirical and computational literary studies (e.g., Fialho 2019; Loi et al. 2023; Miall and Kuiken 1994). One effect of the immense success of the World Wide Web and softwarization and digitization of societies and their cultures (Berry 2014; Manovich 2013) is the availability of large collections of online book reviews and digital full texts from novels published as ePubs. This allows us to apply NLP techniques and corpus statistics to get empirical data on the relations between text and reader that until now could only be theorized or anecdotally evidenced. At the same time, we should acknowledge that it is no panacea for the problem of empirical observations in literary studies. Not just because of the inherent biases (Gitelman 2013; Prescott 2023; Rawson and Muñoz 2016), or the almost complete lack of demographic and social signals in the data, but also because of the difficulties still involved in establishing which concrete signal in novels relates to which



**Figure 1:** Classic rhetorical model (a) and our operationalization of the text–reader relation (b).

type of reaction for which type of reader. This is where we focus our research: We attempt to establish which concrete features of online reviews correlate to which concrete signals in the text of fiction novels.

In a theoretical sense, we are concentrating on the right hand side of the classical rhetorical triangle (see Figure 1a) and operationalize the dynamic between text and reader as another triangular relationship between *impact*, *topic*, and *genre*. With “*impact*” (and the commensurate “*reading impact*”), we designate expressions of reader experiences identified by some evidence-based method (e.g., as reader impact constituents researched by Koolen et al. (2023)). We apply the reader impact model to assign concrete terms to types of reading impact. The concrete text signal that we correlate this impact with are topics mined from a corpus of novels. (As an aside, we note that these topics are not to be confused with themes, motives, or aboutness in a literary studies sense, as we will explain later.) The meta-textual property, *genre*, forms the third measurable aspect of the triangular relationship (see Figure 1b).

Concretely, we link topic models of 18,885 novels in Dutch (original Dutch and translated to Dutch) with the reading impact expressed in 130,751 Dutch online book reviews. We want to know if there is a relationship between aspects of topic in novels, their genre, and the type of impact expressed by readers in their reviews. We extracted expressions for three types of reading impact from the reviews using the previously developed Reading Impact Model for Dutch (Boot and Koolen 2020). The three types of reading impact that we discern are: “General affective impact”, which expresses the overall evaluation and sentiment regarding a novel; “narrative impact”, which relates to aspects of story, plot, and characters; and finally “stylistic impact”, related to writing style and aesthetics.

We expect that topics in fiction are related to genre. As there is no authoritative source for genre of a novel, nor some general academic consensus about what constitutes genre, we make use of the broad genre labels that publishers have assigned to each published book. Analogous to Sobchuk and Šela (2023, 2), who define genre as “a population of texts united by broad thematic similarities”, we clustered these genre labels into a set of nine genres. These thematic similarities might be revealed in a topical analysis, e.g.,

crime novels containing more crime-related topics and romance novels containing more topics related to romance and sex. However, for some genres it might be less obvious whether they are related to topic. For instance, what are the topics one would expect in the broad genre of literary fiction?

It is important to note that, although the name *topic modeling* suggests that what is modeled is *topic*, most topic modeling approaches discern clusters of frequently co-occurring words, regardless of whether they have a topical connection or not (in the classical sense of “aboutness” in library science). Clusters of words may also reveal a different type of connection, e.g., words from a particular stylistic register. In that sense, genres with less clear thematic similarities may be associated with certain stylistic registers, or any other clustering of vocabulary. Different genres may also attract different types of readers and therefore different types of reviewers, who use different terminology and pay attention to different aspects of novels. It is also plausible that the language and topic of a novel influences how readers write about them in reviews. A novel written in a particularly striking poetic style may consciously or subconsciously lead readers to adopt some of its poetic aspects and register in how they write about their reading experiences. Similarly, topics in novels may be associated with what reviewers choose to mention, again, consciously or subconsciously. A novel on the atrocities of war or on the pain of losing a loved one may lead a reviewer to mention feeling sympathy or sadness during reading, while a story about friendship and betrayal might prompt reviewers to describe their anger at the actions of one of the characters.

Thus, it is clear that the relationship between the three elements – topic, genre, and impact – is complex and reciprocal, as expressed in [Figure 1b](#). Our challenge, of course, is to computationally investigate and understand this relationship utilizing the large number of full-text novels from different genres and corpora of hundreds of thousands of reviews. We subdivide this overarching aim into several more concrete research questions, namely:

- How are topic and impact related to each other? Do books with certain topics lead to more impact expressed in book reviews? Do different topics lead to different types of impact?
- How are genre and impact related to each other? Do books of different genres lead to different types of impact? Do reviews of different genres use different vocabulary for expressing the same types of impact?
- How are topic and genre related to each other? Are certain topics more likely in some genres than in others?

This paper makes three main contributions to our ongoing research. The first is that it contributes to our understanding of the reading impact model and through it of the language of reading impact. We formalize the ability to tell genres apart using the *keyness* of impact terms. Thus, we now have quantitative support to argue that certain impact terms are strongly connected to certain genres and less to others. Second, we find that the topics from novels can be clustered into broader themes that lead to distinct thematic profiles per genre. There is a clear relation between impact terms and genre, but not between impact terms and topic or theme. In the discussion at the end, we elaborate on this and provide possible explanations for this finding. The third contribution is

the insight that the key impact terms per genre give an indication of the motivation of readers to read a book and how the reading experience relates to their expectations.

## 2. Background

We are interested in what kind of impression novels leave with their readers. Can we measure this so-called “impact” and how does it relate to features of the actual novel texts? Several studies have tried to link success or popularity of texts to features of those texts. Some studies have related pace, in the sense of how much distance the same length of texts covers in a semantic space, to success; finding that success correlates with higher pacing of narrative (Laurino Dos Santos and Berger 2022; Toubia et al. 2021). It has been argued that songs whose lyrics deviate from a genre’s usual pattern tend to be more popular (Berger and Packard 2018). Other work relating topic models to surveyed ratings of literariness suggests the same for fiction novels (Cranenburgh et al. 2019). Moreira et al. (2023, 32) apply “sentiment arc features [...] and semantic profiling” with some success to predict ratings on Goodreads. Taking the number of Gutenberg downloads as a proxy for success, Ashok et al. (2013) reach 84% accuracy in predicting popularity based on learning low-level stylistic features of the text of novels. Zundert et al. (2018) use sales numbers as a proxy for popularity in a machine learning attempt to predict success, concluding that the theme of masculinity is at least one major driver of successful fiction.

Common to all these studies is that they target some proxy of success or popularity: Goodreads ratings, sales numbers, download statistics, and so forth. However, to our knowledge no research has tried to link concrete features of fiction narratives to textual features of reviews from readers. We seek to uncover if there is such a relation and if it may be meaningful from a literary research perspective. In our present study, we apply a heuristic model for impact features (Boot and Koolen 2020) to a corpus of 600,000+ reader reviews mined from several online review platforms. We attempt to relate collocations of impact related terms to genre. Advancing previous research on genre and topic models (Zundert et al. 2022), our contribution in this paper is to examine how collocated impact terms relate to genre and genre to topic models of novels, thus offering a first insight into the relation between topics (understood in terms of topic model) and reader reported impact measures. Such work needs to take into account the plethora of problems that surround the application of topic models to downstream tasks. This concerns topics content wise, which is to say that topic models in contrast to their name do not often express much topical information. Rather they may be connected to meta-textual features, such as author (Thompson and Mimno 2018), genre (Schöch 2017), or structural elements in texts (Uglanova and Gius 2020).

Our current contribution leans more to the side of data exploration than to the side of offering assertive generalizations. We are interested in empirically quantifying the impact that the text of novels has on readers. Any operationalization of this research aim necessarily involves many narrowing choices and, at least initially, the audacious naivety to ignore the stupefying complexity of social mechanisms to which readers are susceptible and thus the mass of confounding text-external factors that also drive reader impact. In our setup, we assume that there are at least some textual features,

such as style, narrative pace, plot, character likability, that may be measured and that can be related to reader impact. We further assume that book reviews scraped from online platforms do serve as a somewhat reliable gauge to measure reader impact. We make these cautionary statements not just *pro forma*, but because we know that our information is selective, biased, and skewed. Thanks to the stalwart experts of the Dutch National Library, we do have for our analysis the full text of 18,885 novels in Dutch (both translated and of Dutch origin). We also have 634,614 online reviews, gathered by scraping platforms such as Goodreads, Hebban<sup>1</sup>, and so forth. This corpus is biased. Romance novels comprise only about 3% of the corpus of full texts. This is in stark contrast to its undisputed popularity (see Regis (2003, xi): “In the last year of the twentieth century, 55.9% of mass-market and trade paperbacks sold in North America were romance novels”). If our book corpus is skewed, our review data is even more so: Only 1% of the reviews pertain to novels in the romance genre. Obviously, we attempt to balance our data with respect to genre and other properties for analysis. Yet we should remind ourselves of the limited representativeness of our data, which necessitates modesty as to generalizing results. Hence, what follows is more offered as data exploration than as pontification of strong relations.

### 3. Data and Method

Our corpus of 18,885 books consists of mostly fiction novels and some *non-fiction* books in the Dutch language (both originally Dutch and translated). The review corpus boasts 634,614 Dutch book reviews. Obviously, we do not have reviews for each book, nor does the set of books fully cover the collection of reviews, but we have upward of 10,000 books with at least one review.

#### 3.1 Preprocessing

Both – books and reviews – are parsed with Trankit (Nguyen et al. 2021). Reading impact is extracted from the reviews using the Dutch Reading Impact Model (DRIM) (Boot and Koolen 2020).

**Topic modeling** For topic modeling of the novels, we use Top2Vec (Angelov 2020), and created a model with whole books as documents. We apply multiple filters to select terms that signal a topic. Following the advice from previous work (Sobchuk and Šela 2023; Uglanova and Gius 2020; Zundert et al. 2022), we focus on content words, and select only nouns, verbs, adjectives, and adverbs, and remove any person names identified by the Trankit NER tagger. Our assumption is that person names have little to no relationship with topic, but are strong differentiating terms that tend to cluster parts of books and book series with recurring characters. Names of locations can have a similar effect, but, at least where the setting reflects the real world, we argue that this setting aspect of stories is more meaningfully related to topic. The book corpus contains 1,922,833,614 tokens, including all punctuation and stop words. After filtering for person and location names, 826,226,855 tokens remain.

1. See: <https://www.hebban.nl/>.

The next filter is a frequency filter. We remove terms that occur in fewer than 1% of documents or in more than 50% of documents. This leaves 190,607,470 tokens which is 23% of all content words and just under 10% of the total number of tokens<sup>2</sup>. Books have a mean (median) number of 42,959 (37,940) *content* tokens. The number of tokens is a Poisson distribution, therefore left-skewed, with 68% (corresponding to data within 1 standard deviation from the mean) of all books having between 17,509 and 63,418 tokens. This shows that the books have a high variation in length, but the majority of the books have a length within a single order of magnitude. After filtering on document frequency, the mean (median) number of tokens is 9,979 (8,325), with 68% having between 3,847 and 14,992 tokens.

**Reading impact modeling** The DRIM is a rule-based model and works at the level of sentences. It has 275 rules relating to impact in four categories: *Affect*, *Aesthetic* and *Narrative* impact, and *Reflection*. Both *Aesthetic* and *Narrative* impact are sub-categories of *Affect*, so rules that identify expressions of the sub-categories are also considered expressions of *Affect* (Boot and Koolen 2020), but expressions of *Affect* are not necessarily counted as one of the subcategories. The rules for *Reflection* were not validated (see Boot and Koolen 2020), so we exclude *Reflection* from our analysis. For our analysis of topic, we expect that *Narrative* is the most directly related category, but we also include general *Affect* in our analysis. Expressions identified by the model consist of at least an impact word or phrase, such as “spannend” (*suspenseful*<sup>3</sup>). However, many rules require that there is also a book aspect term. For instance, the evaluative word “goed” (*good*) by itself can refer to anything. To be considered part of an impact expression, it must co-occur in one sentence with a word in one of the book aspect categories, e.g. a style-related word like “geschreven” (*written*) to be an expression of *Aesthetic* impact, or a narrative-related word like “verhaal” (*story*) or “plot” to be an expression of *Narrative* impact.

The DRIM identified 2,089,576 expressions of impact in the full review dataset. To identify the key impact terms per genre, we use the full review dataset with all of the approximately 2,1 Mio. impact expressions. To make a clearer distinction between impact expressions of generic affect and affect specific to narrative or aesthetics, we consider as *Affect* only those expressions that are not also categorized as *Narrative* or *Aesthetic*. Of the 2,089,576 expressions, there are 667,672 expressions for *Aesthetic* impact, 690,184 for *Narrative* impact and 731,720 for generic *Affect*.

### 3.2 Connecting Books and Reviews

A crucial step in relating topics in fiction to reading impact expressed in reviews is to connect the books to their corresponding reviews. For this, we rely mostly on the ISBN<sup>4</sup> and the author and the book title. Note that a particular work may be connected to multiple ISBNs, for instance when reprints or new editions are produced for the same work with a different ISBN. Many mappings between reviews and books, and between

2. Experiments with using different frequency ranges for filtering suggests that the topic modeling process is relatively insensitive with regards to the upper limit. I.e., using 50%, 30%, or 10% results in roughly equal numbers of topics that show the same relationship with book genre (see subsubsection 4.1.1 and the following notebook: [https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/notebooks/topic\\_and\\_genre.ipynb](https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/notebooks/topic_and_genre.ipynb)).

3. For all Dutch terms we will consistently provide English translation in italics between parentheses.

4. International Standard Book Number, see: <https://en.wikipedia.org/wiki/ISBN>.

multiple ISBNs of the same work were already made by Boot (2017) and Koolen et al. (2020), for the *Online Dutch Book Response* (ODBR) dataset of 472,810 reviews. We added around 160,000 reviews from Hebban to the ODBR set. To find ISBNs that refer to the same work, we first queried all ISBNs found in reviews using the SRU<sup>5</sup> service of the National Library of the Netherlands. This SRU service gives access to the combined catalog of Dutch libraries and in many cases links multiple editions of the same work with different ISBNs. Using author and title, we resolved another number of duplicated works with different ISBNs. We then mapped all ISBNs of the same work to a unique work ID and linked the reviews via the ISBNs they mention to these work IDs. There are 125,542 distinct works reviewed by the reviews in our dataset. Of the 18,885 books for which we have eBooks, there are 10,056 books with at least one review in our data set. Altogether, these 10,056 unique works are linked to 130,751 reviews.

### 3.3 Connecting Impact and Topic Data

Our goal was to have a comprehensive mapping of the most relevant topics of works to their reviews, the latter analyzed via the DRIM. To create this dataset, we needed to connect the expressions of impact to the topics in our book dataset. To do so, we took the top five dominant topics of each book<sup>6</sup> and linked those topics to the impact expressions in the reviews of the books for that topic. This resulted in a dataset in which each entry links specific reviews to the top five dominant topics for each book.

The Top2Vec model gave us a total of 228 topics. We attempted to label each topic with a distinct content label, but found that many topics are thematically very similar, capturing many of the same elements. Therefore, we manually assigned each topic to one or more of 19 broader themes: 1. *geography & setting*, 2. *behaviors/feelings*, 3. *culture*, 4. *crime*, 5. *history*, 6. *religion, spirituality & philosophy*, 7. *supernatural, fantasy & sci-fi*, 8. *war*, 9. *society*, 10. *city & travel*, 11. *romance & sex*, 12. *medicine/health*, 13. *wildlife/nature*, 14. *economy & work*, 15. *lifestyle & sport*, 16. *politics*, 17. *family*, 18. *science*, 19. *other*. We provide the number of topics grouped per theme in Figure 2<sup>7</sup>.

We provide the full list of topics, themes, and their respective words in our code repository<sup>8</sup>.

### 3.4 Book Genre Information

For genre information about books, we use the Dutch NUR<sup>9</sup> classification codes assigned by publishers. As NUR was designed as a marketing instrument to determine where books are shelved in bookshops, publishers can choose codes based not only on the perceived genre of a book but also on marketing strategies related to where they want a book to be shelved to find the biggest audience. Some NUR codes refer to the same or

5. Search and Retrieval by URL, see: [https://en.wikipedia.org/wiki/Search/Retrieve\\_via\\_URL](https://en.wikipedia.org/wiki/Search/Retrieve_via_URL).

6. Top2Vec creates topics by clustering the document vectors and taking the centroid of each cluster as the topic vector. We computed the cosine similarity between the document vector (representing the book) and the topic vectors, and selected the top five closest (i.e., most similar) topics to each book.

7. Note that in this paper “theme” should not be taken to coincide with the literary studies sense of theme. Rather we use the term “theme” to clearly distinguish between the topics as identified by Top2Vec and their clustering as done by us.

8. See: [https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/data/topic\\_labels.tsv](https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/data/topic_labels.tsv).

9. NUR stands for Nederlandse Uniforme Rubrieksindeling or Dutch Uniform Categories classification.

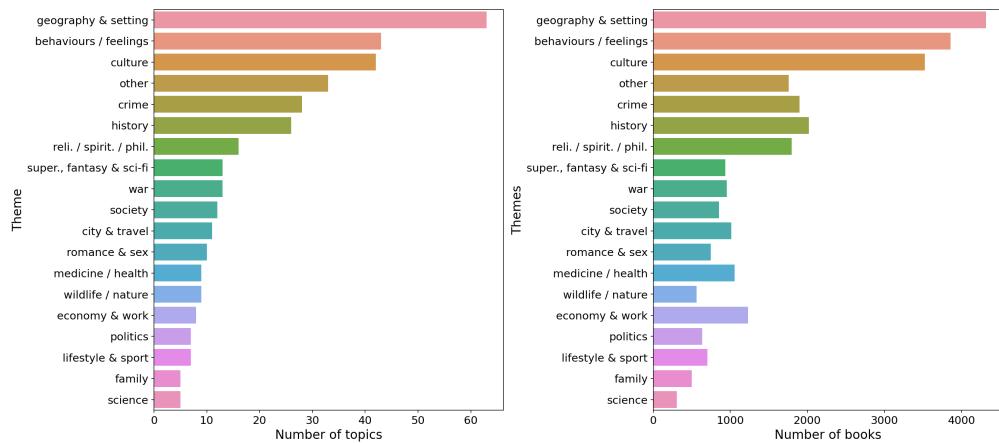


Figure 2: The number of topics and books per theme.

very similar genres. E.g., codes 300, 301, and 302 refer to *general literary fiction*, *Dutch literary fiction*, and *translated literary fiction*, respectively, which we group together under *Literary fiction*. Similarly, we group codes 313, 330, 331, 332, and 339 under *Suspense* novels, as they all refer to types of suspense, i.e., *pocket suspense*, *general suspense novels*, *detective novels*, and *thrillers*, respectively. In total, we select 19 different NUR codes and map them to 9 genres. All remaining NUR codes in the fiction range (300-350) we map to *Other fiction* and the rest to *Non-fiction*. The full mapping is provided in Appendix A.

### 3.5 Keyness Analysis on Impact Terms

The goal of this analysis is to determine (i) *which* words readers use in their reviews to describe the impact of a particular book and (ii) how *characteristic* these words are for a particular genre, compared to another genre. A good candidate to measure both (i) and (ii) is keyword analysis or keyness (Dunning 1994; Gabrielatos 2018; Paquot and Bestgen 2009).

There is ample literature comparing different keyness measures (Culpeper and Demmen 2015; Du et al. 2022; Dunning 1994; Gabrielatos 2018; Lijffijt et al. 2016) and finding that no single measure is perfect. A commonly used measure is  $G^2$ , which identifies *key terms* that occur statistically significantly more or less often in a target corpus (the reviews for a particular genre) compared to a reference corpus (reviews for one or more other genres).

Lijffijt et al. (2016) showed that Log-Likelihood Ratio ( $G^2$ , Dunning 1994) and several other frequency-based bag-of-words keyness measures suffer from excessively high confidence in their estimates because these measures assume samples to be statistically independent, but words in a text are not independent of each other. Du et al. (2022) compare frequency-based and dispersion-based measures for a downstream task (text classification) to show that for identifying key terms in a sub-corpus compared to the rest of the corpus, dispersion-based measures are more effective.

To compare the dispersion of a word or phrase in a target corpus to its dispersion in a reference corpus, Du et al. (2021) introduce *Eta*, which is a variant of the *Zeta* measure by Burrows (2006).

They find that *Eta* (Du et al. 2021) and *Zeta* (Burrows 2006) are among the most effective measures. Both *Eta* and *Zeta* compare document proportions of keywords. The former uses *Deviation of Proportions* (*DP*) (Gries 2008) which computes two sets of proportions. The first are the proportions that the lengths of documents represent with respect to the total number of words in a corpus (e.g., the set of reviews for books of a specific genre) as an expected distribution of the proportions of keywords. The second is the set of observed proportions of a keyword across a corpus with respect to the total corpus frequency of that keyword. There are two problems with using *DP* for keyness of impact terms. The first is that some impact terms do not occur in any of the reviews of a specific genre. In such cases, the observed proportions are not properly defined (a proportion of zero is not well-defined), so *DP* cannot be computed. The second is that the frequency distribution of impact terms in reviews is extremely skewed (84% of all impact terms in reviews have a frequency of 1, while 13% occur twice and the remaining 3% occur three or four times). Although longer reviews have a higher *a priori* probability of containing a specific impact term than shorter reviews, the frequency distribution of individual impact terms behaves more like a binomial distribution, so length-based proportions are not an appropriate measure of keyness.

Because of this, we instead measure dispersion using *document frequencies* (the number of reviews for a book genre in which an impact term occurs) to compute the *document proportion* (the fraction of reviews for a book genre in which an impact term occurs at least once). This gives the document proportion  $docP(t, G)$  per impact term  $t$  and genre  $G$ , with the absolute difference *Zeta* between two genres defined as:

$$Zeta(t, G_1, G_2) = abs(docP(t, G_1) - docP(t, G_2)).$$

To illustrate this approach, we compare the document proportions per genre of the impact terms “stijl” (*style*) and “schrijfstijl” (*writing style*). The former has the highest document proportion for reviews of *Literary fiction* (occurring in 3.7% of the reviews) and least in those of *Non-fiction* (1.2%), resulting in  $Zeta = 0.037 - 0.012 = 0.025$ . The latter is most common in reviews of *Romance* (14.6%) and least common in those of *Non-fiction* (2.0%), giving  $Zeta = 0.146 - 0.02 = 0.126$ .

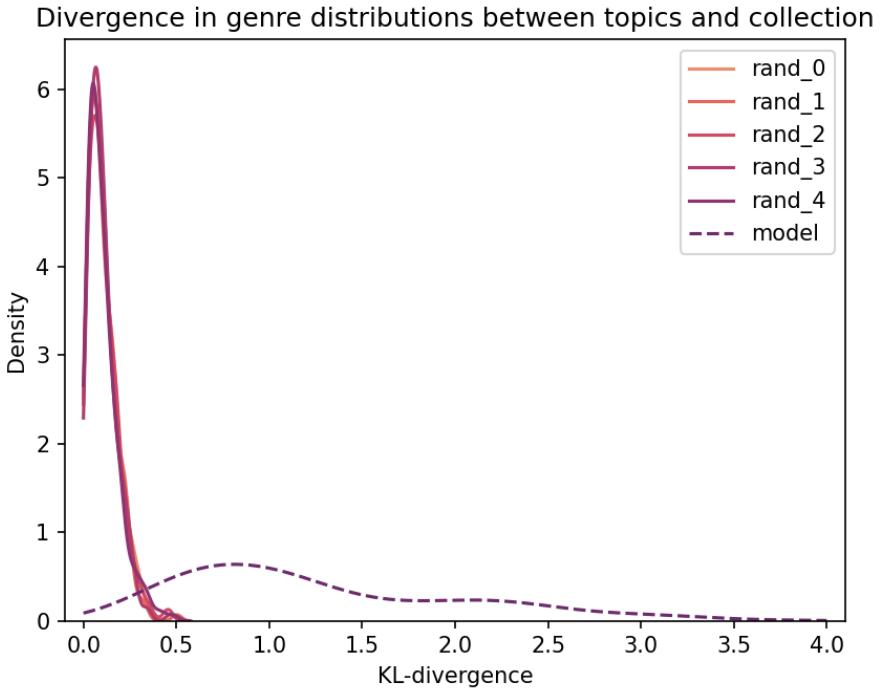
## 4. Results

### 4.1 Topic and Genre

Zundert et al. (2022) found that the topics identified with Top2Vec are strongly associated with genre as identified by publishers. Similarly, Sobchuk and Šešla (2023) find that Doc2Vec – which is used by Top2Vec to embed the documents in the latent semantic space in which topic vectors are identified – is more effective at clustering books by genre than LDA (Blei et al. 2003).

#### 4.1.1 Genre Distribution per Topic

To extend the findings of Zundert et al. (2022), we first quantitatively demonstrate that there is a relationship between topic and genre. Each topic is associated with a number of books and thereby with the same number of genre labels. From eyeballing the distribution of genre labels per topic, it seems that for most topics, the vast majority



**Figure 3:** The KL divergence between the genre distribution per topic and that of the collection for the topic model as well as for five random shuffles of the genre labels using the same books per topic.

of books in that topic belong to a single genre. But the genre distribution of the entire collection is also highly skewed, with a few very large genres and many much smaller genres. So perhaps the skew in most topics resembles the skew of the genre distribution of the collection.

To measure how much the genre distribution per topic deviates from that of the collection, we compute the Kullback-Leibler divergence (KL divergence) between the two distributions.<sup>10</sup> This gives a set of 228 deviations from the collection distribution.

But whether these deviations are small or large is difficult to read from the numbers themselves. For that, we should compare them against a random shuffling of the book genres across books (while keeping the books assigned per topic stable). For large topics (with many books), a random shuffling should have a genre distribution close to that of the collection. For small clusters, the divergence will tend to be higher.

We create five alternative clusterings with books randomly assigned to topics with the same topic size distribution as established by the topic model. The distribution of the 228 KL divergence scores per model (five random and one topic model) are shown in Figure 3. The five random models have almost identical distributions concentrated around 0.1 with a standard deviation of around 0.075 and a max. of around 0.5. The genre distribution of the topic model is very different, with a median score of 1.06 and more than 75% of all scores above 0.68. From this quantitative analysis, it is clear that there is a strong relationship between topic and genre.

10. The KL divergence measures the statistical distance between two distributions, that is, how statistically different they are with respect to each other.

	genre2	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	ooe
genre1		Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA
<b>Children_fiction</b>	14.50	1.08	0.35	0.37	0.29	0.41	0.28	0.07	0.06	0.32	3.52	
<b>Fantasy_fiction</b>	1.08	15.91	0.20	0.23	0.12	0.02	0.07	0.02	0.06	0.24	1.42	
<b>Historical_fiction</b>	0.35	0.20	24.23	1.34	0.26	0.10	0.16	0.98	0.50	0.22	0.81	
<b>Literary_fiction</b>	0.37	0.23	1.34	2.45	1.23	0.21	1.14	0.32	0.81	0.46	0.90	
<b>Literary_thriller</b>	0.29	0.12	0.26	1.23	6.98	0.06	0.34	0.30	0.45	2.17	0.97	
<b>Non-fiction</b>	0.41	0.02	0.10	0.21	0.06	3.23	0.78	0.05	0.03	0.07	0.10	
<b>Other_fiction</b>	0.28	0.07	0.16	1.14	0.34	0.78	3.39	0.87	0.42	0.19	0.46	
<b>Regional_fiction</b>	0.07	0.02	0.98	0.32	0.30	0.05	0.87	13.87	0.44	0.11	0.14	
<b>Romanticism</b>	0.06	0.06	0.50	0.81	0.45	0.03	0.42	0.44	15.39	0.41	0.45	
<b>Suspense</b>	0.32	0.24	0.22	0.46	2.17	0.07	0.19	0.11	0.41	3.32	0.72	
<b>Young_adult</b>	3.52	1.42	0.81	0.90	0.97	0.10	0.46	0.14	0.45	0.72	8.23	

**Figure 4:** Observed over Expected ratio (OoE) of genre co-occurrences as observed in the 228 topics compared to the expected co-occurrences of randomly shuffling the books over 228 clusters of the same size.

We can use the same random shuffling to get more insight into how topics cluster genres. For that, we compute the *observed* co-occurrence of pairs of genres by iterating over all pairs of books in each topic and counting the co-occurrence of their respective genres and divide that by the *expected* co-occurrence of pairs of genres when the books are randomly shuffled. For the *expected* co-occurrence, each shuffling gives different counts, so we repeat the random shuffling 100 times and take the mean number of co-occurrences per pair of genres as the expectation. The Observed over Expected (OoE) ratio is shown in Figure 4. An OoE ratio of 1 means that two genres are co-occurring no more in the topics than is expected when there is no relationship between topic and genre. Scores higher than 1 mean genres are more likely to co-occur than chance (topically, they are similar to each other) and lower than 1 that they are *less* likely to co-occur (topically, they are dissimilar to each other). The numbers on the diagonal are the highest per row and column, meaning that books of each genre are more likely to end up in topics with other books of the same genre than with books of a different genre.

We make a few more observations. First, some genres are very dissimilar from others. Most of the compared OoE scores for *Fantasy* with other genres are well below 1.0. It is topically only slightly similar to *Young adult*. Second, some genres are topically similar to each other. *Children's fiction* and *Young adult* have an OoE of 3.52, while the OoE of *Literary thriller* and *Suspense* is 2.17. These are topical connections that are not surprising. Third, *Literary fiction* is topically somewhat similar to *Literary thrillers* (OoE of 1.23), but dissimilar to *Suspense* (0.46). Even though *Literary thriller* is similar to *Suspense*, it has a topical connection to other *Literary fiction* that *Suspense* does not have. In addition, while NUR codes are mostly a marketing instrument, their distinction between *Literary thrillers* and other *Suspense* novels relates to a topical distinction as well. Finally, fourth, the numbers on the diagonal vary strongly, with *Historical fiction* novels being much more likely to be topically clustered with other historical novels than with novels of other genres (OoE is 24.23), while for *Literary fiction* (2.45) and *Non-fiction* (3.23) this is much less likely. This may be partly due to the fact that the latter two are the largest genres in the collection and therefore have a high *a priori* probability to end up in topics

with books of other genres, but we speculate that it may also be due to the fact that these two genres do not have a clear topic profile (whereby we stress that topic here is interpreted as sharing vocabulary, because the Doc2Vec embedding space is based on word tokens).

#### 4.1.2 Thematic Distribution per Genre

Next, we perform a qualitative analysis of the topics and their relationship to genre, via the identified themes described in [subsection 3.3](#).

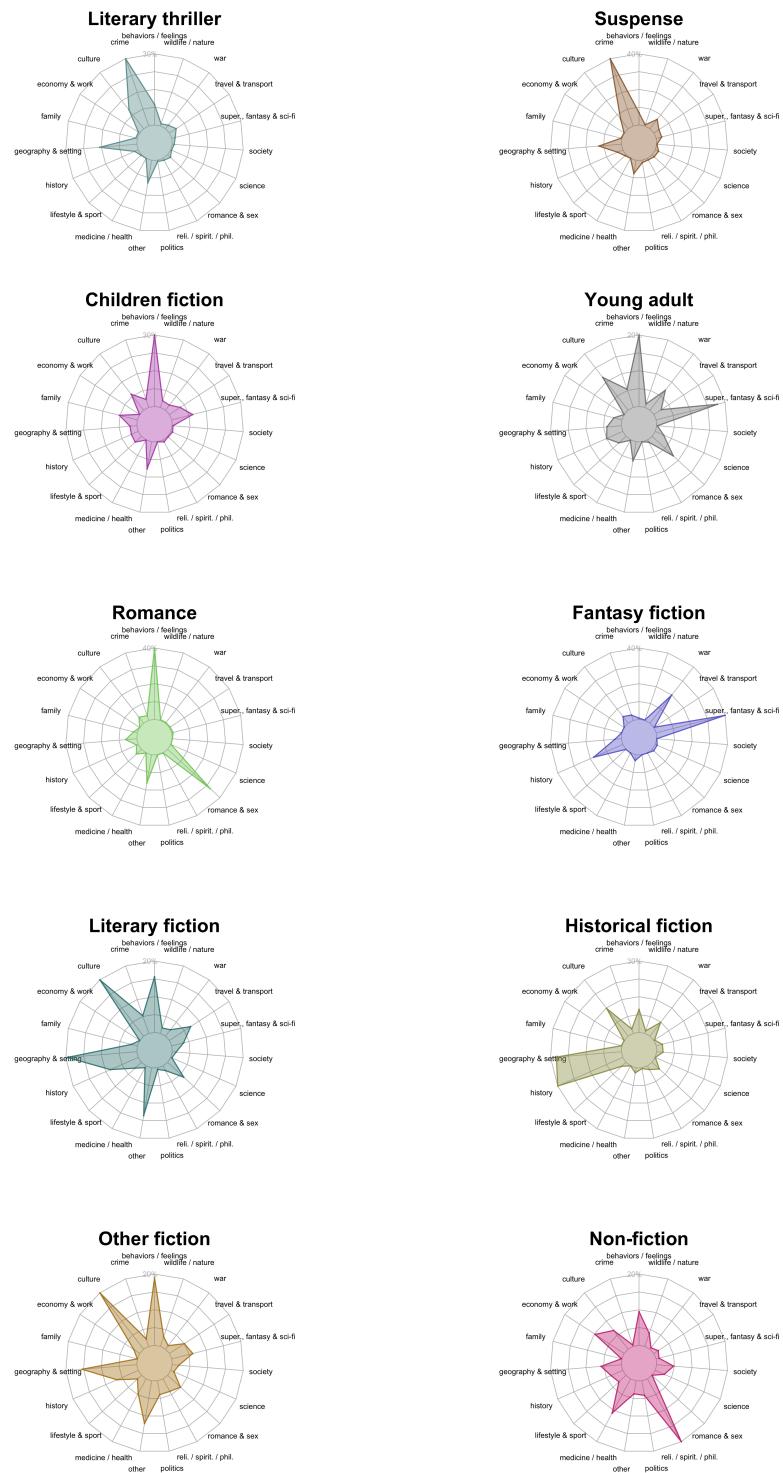
The distribution of topic themes per genre is shown in [Figure 5](#) in the form of radar plots. The genres show distinct thematic profiles. *Literary fiction* scores high on the themes of *culture, geography & setting* and *behaviors/feelings* which is perhaps not surprising. *Non-fiction* scores high on *religion, spirituality & philosophy, medicine/health, economy & work*, and *behaviors/feelings* which are themes that few fiction genres score high on.

In *Children's fiction*, there is relatively little use of the geographical aspect of setting, especially compared to other fiction genres. That is, it seems that children's novels make little explicit reference to geographical places. They score high on *behaviors/feelings* and moderately high on *culture, family* and *supernatural, fantasy & sci-fi*. The main difference between *Children's fiction* and *Young adult* is that the latter scores higher on *supernatural, fantasy & sci-fi*. For the former, *Young adult* strongly overlaps with *Fantasy* novels. *Young adult* also adds in a bit of *romance & sex*. These observations suggest that *Children's fiction* and *Young adult* by and large treat the same themes, but against different 'backgrounds'. *Children's fiction* deals with *behaviors/feelings* against a backdrop of *culture* and *family*. *Young adult* does practically the same, but adds *supernatural, fantasy & sci-fi* elements to the story and opens the stage for some romantic behavior.

If one were to hazard a guess about reader development, it would almost seem as if young readers are invited to pre-sort on the major themes of grown-up literature, with *Romance* amplifying the *romance & sex* encountered in *Young adult* books, while *Literary fiction* and *Literary thrillers* amplify motifs of *culture, setting*, and *crime*, and *Fantasy* caters to the interest in the supernatural developed through *Young adult* fiction. Much more research would be needed, however, to substantiate such a pre-sorting effect. In any case, *Romance* scores high on *romance & sex* and has medium scores for *culture* and *geography & setting*, while *Suspense* novels score high on *crime* and have medium scores for *geography & setting* and *war*.

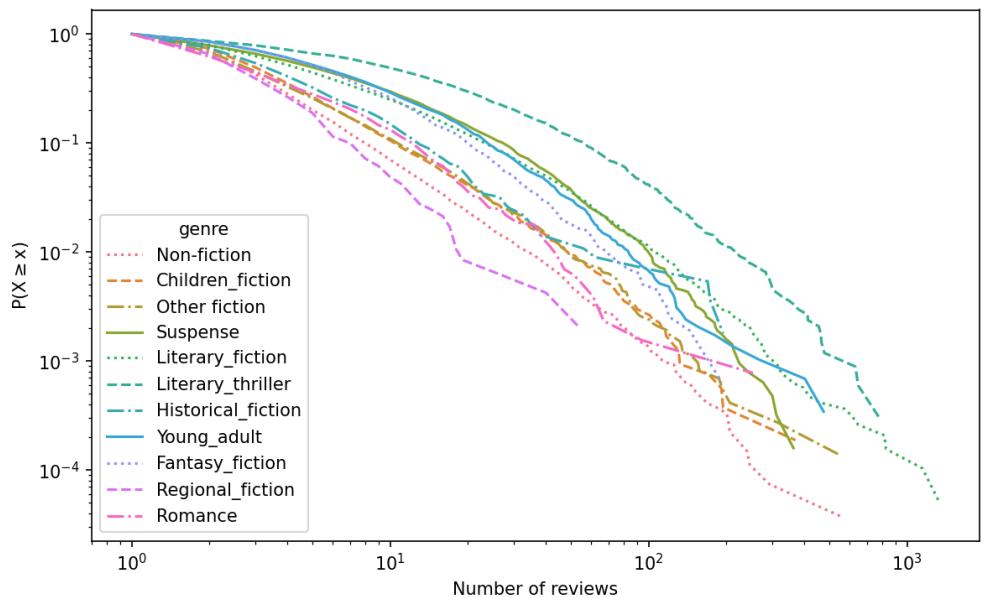
We expect that many of these observations coincide with intuitions of literary researchers. This suggests that the grouping of topics by theme makes sense from a literary analytical perspective. The findings also show where genres overlap and where they differ. For instance, the profile for *Literary fiction* and *Literary thriller* are similar, with the main difference being the much higher prevalence of the *crime* theme in *Literary thrillers*. *Suspense* is similar to *Literary thrillers* in the prevalence of *crime* as theme, but lower scores for *culture* and *geography & setting*.

One of the main findings is that, for the chosen document frequency range of mid-frequency terms, there is a clear connection between topic and genre, with thematic clustering of topics leading to distinct genre profiles, but also to thematic connections between certain genres. None of this will radically transform our understanding of genre



**Figure 5:** Radar plots showing the relative prevalence of themes in six genres, from left to right, top to bottom: *Literary thrillers*, *Suspense*, *Children's fiction* and *Young adult*, *Romance*, *Fantasy fiction*, *Literary fiction*, *Historical fiction*, *Other fiction* and *Non-fiction*.

	Reviewed books	Reviews	Mean reviews/book
Literary fiction	19,288	200,907	10.4
Literary thriller	3,394	77,288	22.8
Young adult	2,919	30,552	10.5
Children fiction	5,348	27,989	5.2
Suspense	6,266	67,990	10.9
Fantasy fiction	1,571	13,739	8.7
Romance	1,291	6,434	5.0
Historical fiction	556	3,463	6.2
Regional fiction	472	1,528	3.2
Other fiction	7,260	37,515	5.2
Non-fiction	26,884	109,158	4.1

**Table 1:** Reviews per genre and mean number of reviews per book per genre.**Figure 6:** The cumulative distribution function of the number of reviews per book, on a log-log scale. The Y-axis shows the probability  $P(X \geq x)$  that a book has at least  $x$  reviews.

and topic, but it prompts the question how different parts of the document frequency distribution relate to different aspects of novels. From authorship attribution research, we know that authorial signal is mainly found in the high-frequency range and our work corroborates earlier findings that topics contain genre-signals in mid-range frequencies (Thompson and Mimno 2018; Zundert et al. 2022).

## 4.2 Impact and Genre

### 4.2.1 Reviews per Genre

With the genre labels, we can count how many books in each genre have reviews in our dataset and how many reviews they have (Table 1). The genre with the highest total number of reviews is *Literary fiction*, with 200,907 reviews in our dataset, followed by *Literary thrillers* and *Suspense* novels. If we consider the number of reviews per book, *Literary thrillers* have the highest mean number of reviews (22.8). However, the distribution of the number of reviews per book is highly skewed, with a single review

per book being the most likely and having more reviews being increasingly unlikely (Koolen et al. 2020). The distributions per genre show some differences, but all are close to a power-law. The cumulative distribution function of the number of reviews per book for the different genres are shown in Figure 6, with on the Y-axis the probability  $P(X \geq x)$  that a book has at least  $x$  reviews.<sup>11</sup>

The curves for some of the genres overlap, which makes them difficult to discern, but there are a few main insights. First, *Regional fiction* and *Non-fiction* have the fastest falling curves, indicating that books in these genres are the least likely to acquire many reviews. Next is a cluster of *Children's fiction*, *Romance*, *Historical fiction*, and *Other fiction*, which tend to get a slightly higher number of reviews. Then there is a cluster of *Suspense*, *Literary fiction*, *Young adult*, and *Fantasy fiction*, which tend to get more reviews than the previous cluster. And finally, clearly above the rest, is the curve of *Literary thrillers*, which tend to get more reviews than books in any other genre.

*Thrillers* are more often reviewed on the platforms that are in the review dataset. *Romance* novels have fewer reviews but are a very popular genre (Regis (2003, 108), see also Darbyshire (2023)). This prompts the question of whether readers of *Regional* and *Romance* novels have less desire to review these novels or review them on different platforms and in different ways. As there seem to be many video reviews of *Romance* novels on TikTok using the tag #BookTok, this would be a valuable resource to add to our investigations. A difference in the number of reviews might be a signal of a difference in impact, but it is also plausible that different genres attract different types of readers who express their impact in different ways linguistically, using different media (e.g., text or video) on different platforms (e.g., GoodReads or TikTok). To that extent, the review dataset may be a biased representation of the impact of books in different genres. Bracketing for a moment the potential skewedness of the number of reviews per genre and taking the number of reviews as a proxy of popularity, it is also interesting to observe that popularity is apparently a commodity that is reaped in orders of magnitude.

#### 4.2.2 Key Impact Terms per Genre

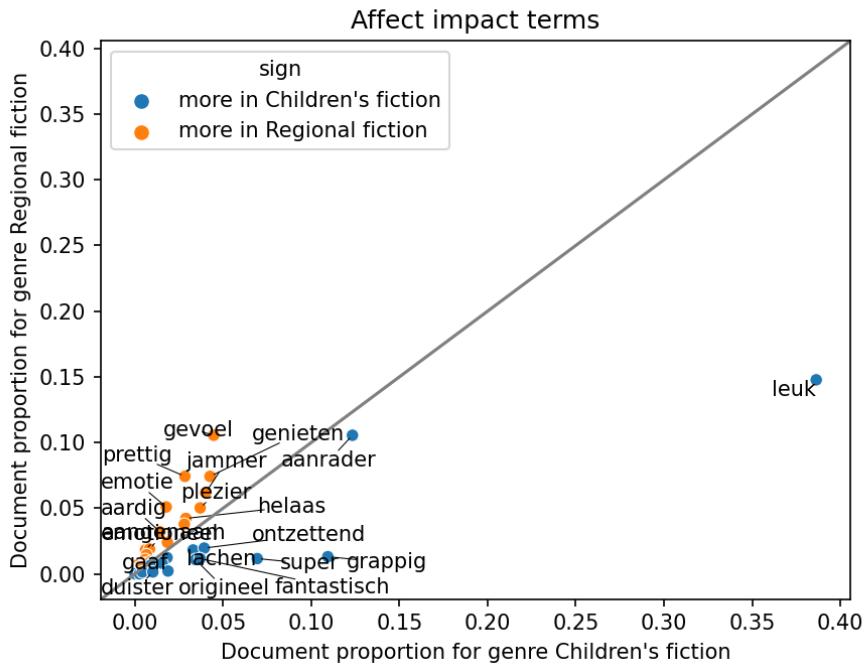
**Correlations between genres** First, we compare genres in terms of their impact terms using the document proportions per impact term. For each pair of genres, we compute the Pearson correlation  $\rho$  between the document proportions of all impact terms. A high positive correlation means that impact terms with a high document proportion in one genre tend to also have a high document proportion in the other genre.

The correlations per impact type are shown in Figure 7. For *Affect* impact terms (the top correlation table), most genre pairs have a near perfect correlation ( $0.8 < \rho < 1.0$ ) and only few pairs have a moderate ( $0.4 < \rho < 0.6$ ) or strong correlation ( $0.6 < \rho < 0.8$ ), notably *Children's fiction* in combination with either *Historical fiction*, *Literary thrillers* and *Suspense*. For *Narrative* impact terms, there are more moderate correlations, with *Non-fiction* standing out as the most distinct genre. This is not surprising, given that (we assume) *Non-fiction* books are least likely to be discussed in terms of narrative. For *Aesthetic* impact terms, there are only four correlations below but close to 0.8, indicating

11. We show the cumulative distribution instead of the plain distribution because it produces smoother curves and better shows the trends.

	Affect												
	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA	Unkn.	
<b>Child. fic</b>	1.000	0.812	0.598	0.765	0.564	0.938	0.866	0.738	0.917	0.617	0.895	0.912	
<b>Fantasy</b>	0.812	1.000	0.875	0.933	0.882	0.873	0.923	0.880	0.893	0.912	0.962	0.953	
<b>Hist. fic</b>	0.598	0.875	1.000	0.938	0.886	0.770	0.896	0.942	0.800	0.908	0.831	0.845	
<b>Lit. fic</b>	0.765	0.933	0.938	1.000	0.873	0.853	0.976	0.944	0.910	0.900	0.931	0.953	
<b>Lit. thrill</b>	0.564	0.882	0.886	0.873	1.000	0.694	0.802	0.825	0.738	0.996	0.806	0.812	
<b>Non-fic</b>	0.938	0.873	0.770	0.853	0.694	1.000	0.935	0.827	0.905	0.743	0.913	0.943	
<b>Oth. fic</b>	0.866	0.923	0.896	0.976	0.802	0.935	1.000	0.939	0.953	0.842	0.952	0.978	
<b>Reg. fic</b>	0.738	0.880	0.942	0.944	0.825	0.827	0.939	1.000	0.918	0.858	0.892	0.906	
<b>Romance</b>	0.917	0.893	0.800	0.910	0.738	0.905	0.953	0.918	1.000	0.781	0.962	0.963	
<b>Suspense</b>	0.617	0.912	0.908	0.900	0.996	0.743	0.842	0.858	0.781	1.000	0.846	0.851	
<b>YA</b>	0.895	0.962	0.831	0.931	0.806	0.913	0.952	0.892	0.962	0.846	1.000	0.979	
<b>Unkn.</b>	0.912	0.953	0.845	0.953	0.812	0.943	0.978	0.906	0.963	0.851	0.979	1.000	
	Narrative												
	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA	Unkn.	
<b>Child. fic</b>	1.000	0.943	0.829	0.774	0.933	0.498	0.685	0.693	0.583	0.929	0.954	0.934	
<b>Fantasy</b>	0.943	1.000	0.886	0.805	0.970	0.415	0.661	0.722	0.657	0.970	0.979	0.953	
<b>Hist. fic</b>	0.829	0.886	1.000	0.908	0.892	0.470	0.777	0.855	0.748	0.896	0.914	0.924	
<b>Lit. fic</b>	0.774	0.805	0.908	1.000	0.803	0.712	0.936	0.888	0.712	0.802	0.852	0.921	
<b>Lit. thrill</b>	0.933	0.970	0.892	0.803	1.000	0.397	0.644	0.691	0.604	0.999	0.969	0.958	
<b>Non-fic</b>	0.498	0.415	0.470	0.712	0.397	1.000	0.885	0.669	0.476	0.391	0.485	0.595	
<b>Oth. fic</b>	0.685	0.661	0.777	0.936	0.644	0.885	1.000	0.868	0.681	0.644	0.729	0.814	
<b>Reg. fic</b>	0.693	0.722	0.855	0.888	0.691	0.669	0.868	1.000	0.860	0.692	0.774	0.798	
<b>Romance</b>	0.583	0.657	0.748	0.712	0.604	0.476	0.681	0.860	1.000	0.614	0.696	0.686	
<b>Suspense</b>	0.929	0.970	0.896	0.802	0.999	0.391	0.644	0.692	0.614	1.000	0.972	0.957	
<b>YA</b>	0.954	0.979	0.914	0.852	0.969	0.485	0.729	0.774	0.696	0.972	1.000	0.973	
<b>Unkn.</b>	0.934	0.953	0.924	0.921	0.958	0.595	0.814	0.798	0.686	0.957	0.973	1.000	
	Aesthetic												
	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA	Unkn.	
<b>Child. fic</b>	1.000	0.936	0.953	0.972	0.832	0.957	0.975	0.961	0.904	0.822	0.949	0.979	
<b>Fantasy</b>	0.936	1.000	0.963	0.972	0.964	0.893	0.973	0.969	0.947	0.960	0.987	0.963	
<b>Hist. fic</b>	0.953	0.963	1.000	0.977	0.897	0.947	0.972	0.983	0.874	0.880	0.963	0.970	
<b>Lit. fic</b>	0.972	0.972	0.977	1.000	0.899	0.939	0.992	0.969	0.909	0.887	0.964	0.993	
<b>Lit. thrill</b>	0.832	0.964	0.897	0.899	1.000	0.779	0.899	0.920	0.915	0.998	0.956	0.885	
<b>Non-fic</b>	0.957	0.893	0.947	0.939	0.779	1.000	0.956	0.926	0.820	0.763	0.895	0.957	
<b>Oth. fic</b>	0.975	0.973	0.972	0.992	0.899	0.956	1.000	0.969	0.929	0.889	0.966	0.993	
<b>Reg. fic</b>	0.961	0.969	0.983	0.969	0.920	0.926	0.969	1.000	0.923	0.907	0.987	0.966	
<b>Romance</b>	0.904	0.947	0.874	0.909	0.915	0.820	0.929	0.923	1.000	0.923	0.953	0.902	
<b>Suspense</b>	0.822	0.960	0.880	0.887	0.998	0.763	0.889	0.907	0.923	1.000	0.950	0.873	
<b>YA</b>	0.949	0.987	0.963	0.964	0.956	0.895	0.966	0.987	0.953	0.950	1.000	0.959	
<b>Unkn.</b>	0.979	0.963	0.970	0.993	0.885	0.957	0.993	0.966	0.902	0.873	0.959	1.000	

**Figure 7:** Pearson correlation in the doc proportion scores of impact terms between pairs of genres, for *Affect* (top), *Narrative* (middle) and *Aesthetic* (bottom).



**Figure 8:** Document proportions of generic *Affect* terms for *Children's fiction* and *Regional fiction*.

that there are few differences in vocabulary between genres. The overwhelming majority of strong and near perfect correlations suggests that, overall, impact across genres is expressed in the same vocabulary.

**Vocabulary differences between genres** Even though the correlations are mostly strong, we can still zoom in on the largest differences in vocabulary usage. For generic *Affect*, *Children's fiction* is most distinctive as it has high score differences with all other genres. The document proportions for generic *Affect* terms of *Children's fiction* and *Regional fiction* are shown in Figure 8. The diagonal line shows where terms have equal proportions in both genres. Reviews of *Children's fiction* seem to use a smaller impact vocabulary – almost all document proportions are close to zero – but much higher proportions for the impact term “leuk” (*fun* or *cool*). This term is used much less in reviews of other genres.

For *Narrative* impact, the biggest summed difference is between *Romance* and *Literary thrillers* (see Figure 9). The main differences are found with a handful of terms, “spannend” (*thrilling, suspenseful*), “spanning” (*suspense*) and “verrassen” (*surprise*) are more common in *Literary thrillers* and “romantisch” (*romantic*) and “heerlijk” (*lovely, wonderful*) are more common in *Romance* novels. These are perhaps somewhat obvious, but show that impact, or at least the language of impact, is related to genre.

For *Aesthetic* impact, the biggest summed difference is between *Romance* and *Historical fiction* (see Figure 10). Here again, the main differences are in a few terms. Reviews of *Historical fiction* more often mention impact terms like “mooi” (*beautiful*), “beschrijven” (*describe*), “beschreven” (*described*), and “prachtig” (*beautiful*). Reviews of *Romance* novels more often mention “schrijfstijl” (*writing style*), “humor” (*humor*), and “luchtig” (*airy*). It seems that for *Historical fiction*, reviewers focus more on descriptions (how evocatively the author describes historical settings, persons, or events), while reviewers

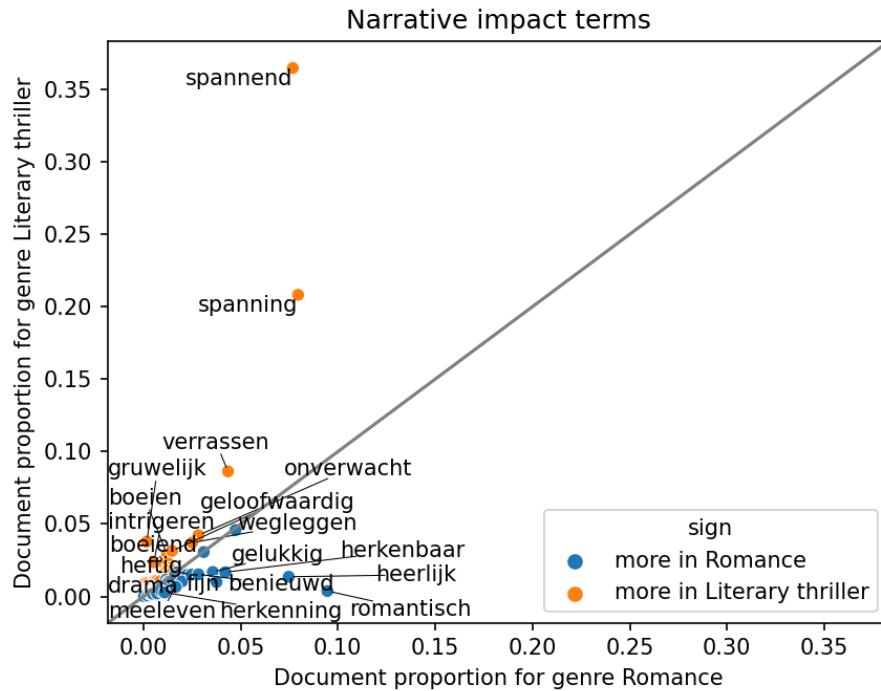


Figure 9: Document proportions of *Narrative impact terms* for *Romance* and *Literary thrillers*.

of *Romance* novels focus more on humor and lightness of style. A close reading of some of the contexts in which “schrijfstijl” is mentioned in *Romance* reviews suggests that reviewers often use it in phrases like “makkelijke schrijfstijl” and “vlotte schrijfstijl” (*a writing style that reads easily or quickly*, respectively).

### 4.3 Impact and Topic

The third link between the three main concepts that are the focus of this paper is between impact and topic.

To study how the use of impact terms differs between reviews of books with different themes – recall, we are talking about theme in the sense of topically grouped clusters of books – we first need to group the reviews by theme. Because themes are based on topics and some themes share the same topics, some reviews are assigned to multiple themes. We calculated Pearson correlations between themes in terms of the document proportions per impact term, just as we did for genre (see Figure 13, Figure 14, and Figure 15 in Appendix C). There are many observations that could be made, but again, we limit ourselves to the most salient ones related to the three largest themes (in number of books). First of all, the vast majority of the correlations are near perfect, suggesting that impact is expressed with similar vocabulary across reviews for books associated with different themes. For *Aesthetic* impact, there are no correlations below 0.8. For *Narrative* impact, the one clearly distinct theme is *medicine/health*, which has no or weak correlations with most of the other themes.

When we zoom in on the document proportions of individual impact terms and compare two genres, we observe the overall similarity but also some specific differences. The

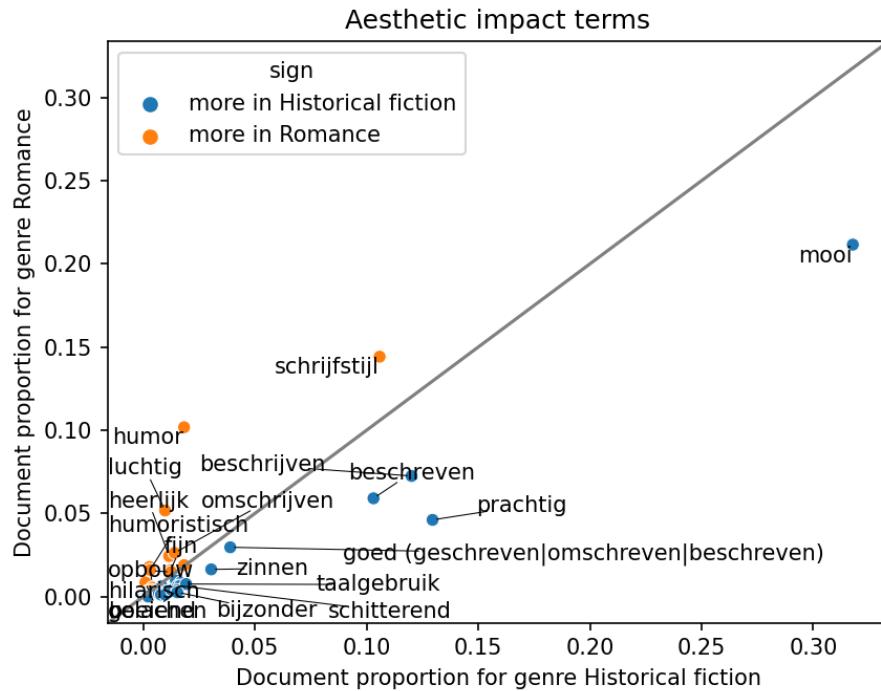


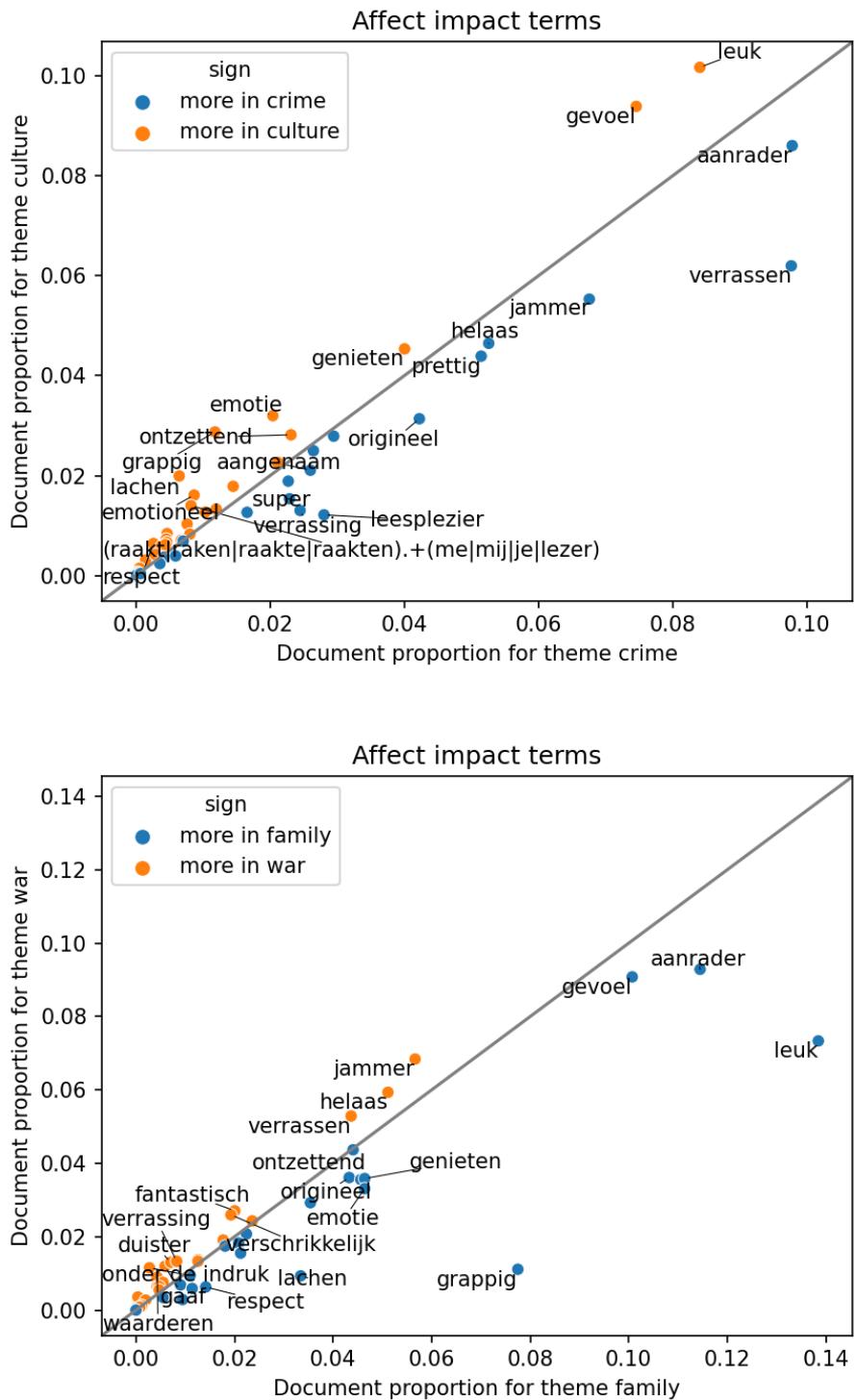
Figure 10: Document proportions of Aesthetic impact terms for *Historical fiction* and *Romance*.

comparative document proportions for general *affect* terms are shown for reviews of books related to themes *crime* and *culture* (top of Figure 11) or *family* and *war* (bottom of Figure 11).

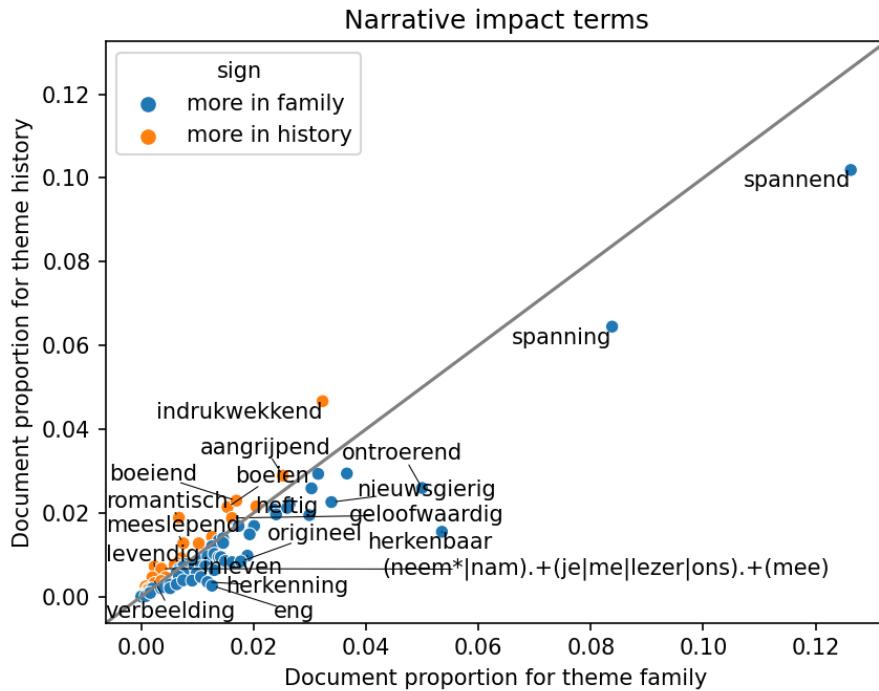
The proportions for *crime* and *culture* are slightly different, but most of the data points are close to the diagonal and the correlation between the sets of proportions is high. Impact terms like “verrasen” (*to surprise*) and “aanrader” (*recommendation*) are used slightly more often in reviews of *crime*-related books, while terms like “gevoel” (*feeling*), “emotie” (*emotion*), and “grappig” (*funny*) are more often used for *culture*-related books. However, the relative differences in proportion are small. For *family* and *history*, we observe larger differences, with *affect* terms like “leuk” (*fun, enjoyable*) and especially “grappig” (*funny*) having much higher document proportions in reviews of *family*-related books than *war*-related books.

For *Narrative* impact terms, the comparative document proportions for reviews related to the themes *family* and *history* are shown in Figure 12. Again, most of the data points are close to the diagonal, showing the similarity in usage of impact terms. However, the biggest difference is that *family* reviewers are more likely to use terms like “herkenbaar” (*recognizable*), “ontroerend” (*touching*), while *history* reviewers more often use “indrukwekkend” (*impressive*), “aangrijpend” (*gripping*), and “boeiend” (*intriguing, fascinating*).

Note that for terms with lower document proportions (i.e., between 0 and 0.1%), the relative differences in proportions can be large, signaling potentially highly statistically significant differences between genres or themes. But, the fact that the proportions are low means that these significant differences are between very rare and extremely rare usage of terms. To illustrate, the *Aesthetic* term “geniaal” (*genius or brilliant*) is six times



**Figure 11:** Document proportions of general Affect terms for the themes *crime* and *culture* (top) and *family* and *war* (bottom).



**Figure 12:** Document proportions of *Narrative impact terms* for the themes *family* and *history*.

more likely to occur in *Young adult* fiction reviews than in reviews of *Children’s fiction*, but “geniaal” is very rare in the former (14 total occurrences, or 0.05% of 29,075 reviews) and extremely rare in the latter (two occurrences, or 0.008% of 25,074 reviews).

Although such large relative differences may give further insight into how genre and theme relate to impact, we want to stress the high overall similarity. It suggests that reviewers use a largely common vocabulary for expressing impact, regardless of the genre or theme of a book. Large relative differences in rare terms are potentially insightful to interpret differences between individual books, authors, or reviewers, but they say little about genres or topics overall.

## 5. Discussion and Conclusion

In this paper, we investigated the relationship between three important concepts in literary studies: genre, topic, and impact. We discuss our findings for each pair of concepts in turn.

**Genre and topic** Our analyses have corroborated earlier findings on the relationship between genre and topic. By clustering topics identified by topic modeling into broader themes and by measuring the prevalence of these themes in the books of specific genres, we find that topics have a strong relation with genres and the genres have distinct thematic profiles. These profiles match existing intuitions about the distribution of themes across genres. Potentially, these profiles can provide additional insight into genre dynamics (e.g., as to what motivates readers to mix-read genres or not), although much of this aspect remains to be examined.

**Genre and impact** The Dutch Reading Impact Model (DRIM by Boot and Koolen (2020)) identifies sets of words that are to some extent related to genre, and by studying the overlap in key impact terms between genres, we find clusters of genres that are similar in how their impact is described. Of course, this is not entirely surprising. For instance, *Suspense* novels and *Literary thrillers* are highly similar in terms of all three types of impact. However, it is much less obvious or intuitive that *Historical fiction*, *Literary fiction* and *Fantasy* have very similar distributions of *Aesthetic* impact terms, nor that *Non-fiction* is distinct from most other genres in terms of *Narrative* impact, apart from *Literary fiction* and *Other fiction*.

It remains unclear for now how we should explain the relationship between impact and genre. Perhaps this relation signals that reviewers develop and copy conventions for writing about books from other reviews they have read, regardless of genre differences. At the same time, we should not ignore the differences that do exist. At an aggregate level, differences may seem small, but small differences in usage across a range of impact terms could still signal a consistent and meaningful difference in impact. Finally, depending on how the reading impact model was developed, this may also be an artifact of how the rules were constructed. For instance, if reviews for a heterogeneous set of books were scanned to identify recurring expressions of impact, it is possible that expressions that are shared across genres stood out and were more likely to be included in the set of rules. Further analysis is required to establish which, if any, of these factors contributes to the relationship between fiction genres and reading impact as expressed in reviews.

**Topic and impact** For the first two pairs of concepts, there were some expectations, e.g., that there is a relation between the *Romance* genre and topics related to the theme of *romance & sex*, or that typical narrative impact terms in reviews of *Young adult* novels overlap with those in reviews of *Fantasy* novels. For the link between topic and impact, we struggled to come up in advance with expectations on how the topics in novels are related to impact. Novels discussing topics such as *war* and its consequences or living with physical or mental illness might lead to more reviews mentioning *Narrative* impact. But honest reflection forces us to admit that the results of topic modeling do not shed much light on how authors deal with topics and how reviewers discuss them. This gap stubbornly persists throughout continued engagement with our data in several papers. Consequently, this should give us pause to reflect on our operationalizations. Although vector models have moved beyond bag-of-words approaches and are becoming increasingly more sophisticated, we have not inched significantly closer to answering the question of which features of novel texts relate to what types of reader impact adequately and satisfactorily from a literary studies perspective.

Our reflections tie in with observations and suggestions made in some recent methodological publications on computational humanities: Bode (2023) argues that humanities researchers applying conventional methods and those embracing computational or data science methods should take a greater and more sincere interest in each other's work. Rather than addressing research questions by stretching either method beyond its limits, researchers ought to investigate how the different methods can reinforce and amplify each other. Pichler and Reiter (2022) argue that operationalizations in computational

linguistics and computational literary studies are currently often poor because we typically fail to express the precise operations that identify the theoretical concept we are trying to observe. Indeed our operationalizations seem underwhelming in the light of literary mechanisms. The reason to label a topic as being *about* war is that it contains words directly and strongly associated with war and emphasizing the physical aspects of it, such as *war*, *soldier*, *bombing*, *battlefield*, *wounded*, etc. But novels that readers would describe as being *about* war might instead focus on more indirect aspects or on aspects that war shares with many other situations, such as dire living conditions or being cut-off from the rest of the world, feeling unsafe and scared, or the sense of helplessness or hopelessness. The problem is not just that words indirectly related to war might lead an annotator to label a topic as being about something other than war. It is also that an author, going by the good practice of “show don’t tell”, can conjure up images that fit these words in almost infinitely many ways that are almost impossible to capture by looking at bags of words. Which means we need infinitely better operationalizations.

## 6. Data Availability

Data used for the research can be found at: <https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact>. It has been archived and is persistently available at <https://doi.org/10.5281/zenodo.13929510>.

## 7. Software Availability

All code created and used in this research has been published at: <https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact>. It has been archived and is persistently available at <https://doi.org/10.5281/zenodo.13929510>.

## 8. Acknowledgements

This project has been supported through generous material and in-kind technical and data-science analytical support from the eScience Center in Amsterdam. We thank the National Library of the Netherlands for providing access to the novels used in this research and for their invaluable technical support.

## 9. Author Contributions

**Marijn Koolen:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft

**Joris J. van Zundert:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Visualization, Writing – review & editing

**Eva Viviani:** Formal analysis, Software, Validation, Visualization

**Carsten Schnober:** Resources, Software

**Willem van Hage:** Methodology, Resources, Software

**Katja Tereshko:** Writing – original draft, Writing – review & editing

## References

- Abrams, Meyer H. (1971). *The Mirror and the Lamp: Romantic Theory and the Critical Tradition*. Oxford University Press.
- Angelov, Dimo (2020). "Top2Vec: Distributed Representations of Topics". In: *arXiv*. [10.48550/arXiv.2008.09470](https://arxiv.org/abs/10.48550/arXiv.2008.09470).
- Ashok, Vikas Ganjigunte, Song Feng, and Yejin Choi (2013). "Success with Style: Using Writing Style to Predict the Success of Novels". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1753–1764. <https://api.semanticscholar.org/CorpusID:7100691> (visited on 07/28/2023).
- Berger, Jonah and Grant Packard (2018). "Are Atypical Things More Popular?" In: *Psychological Science* 29 (7), 1178–1184. [10.1177/0956797618759465](https://doi.org/10.1177/0956797618759465).
- Berry, David M. (2014). *Critical Theory and the Digital*. Critical Theory and Contemporary Society. Bloomsbury Academic.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3 (1), 993–1022. <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf> (visited on 10/03/2024).
- Bode, Katherine (2023). "What's the Matter with Computational Literary Studies?" In: *Critical Inquiry* 49 (4), 507–529. [10.1086/724943](https://doi.org/10.1086/724943).
- Boot, Peter (2017). "A Database of Online Book Response and the Nature of the Literary Thriller". In: *Book of Abstracts of DH 2017*. <https://dh2017.adho.org/abstracts/208/208.pdf> (visited on 10/07/2024).
- Boot, Peter and Marijn Koolen (2020). "Captivating, Splendid or Instructive? Assessing the Impact of Reading in Online Book Reviews". In: *Scientific Study of Literature* 10 (1), 35–63. [10.1075/ssol.20003.boo](https://doi.org/10.1075/ssol.20003.boo).
- Burrows, John (2006). "All the Way through: Testing for Authorship in Different Frequency Strata". In: *Literary and Linguistic Computing* 22 (1), 27–47.
- Cranenburgh, Andreas van, Karina van Dalen-Oskam, and Joris van Zundert (2019). "Vector Space Explorations of Literary Language". In: *Language Resources and Evaluation* 53 (4), 625–650. [10.1007/s10579-018-09442-4](https://doi.org/10.1007/s10579-018-09442-4).
- Culpeper, Jonathan and Jane Demmen (2015). "Keywords". In: *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press, 90–105.
- Darbyshire, Madison (2023). "Hot Stuff: Why Readers Fell in Love with Romance Novels". In: *Financial Times*. <https://www.ft.com/content/0001f781-4927-4780-b46c-3a9f15dfffe78> (visited on 10/03/2024).
- Du, Keli, Julia Dudar, Cora Rok, and Christof Schöch (2021). "Zeta & Eta: An Exploration and Evaluation of Two Dispersion-based Measures of Distinctiveness". In: *Proceedings of Computational Humanities Research*, 181–194. [https://ceur-ws.org/Vol-2989/shrt\\_paper11.pdf](https://ceur-ws.org/Vol-2989/shrt_paper11.pdf) (visited on 10/03/2024).
- Du, Keli, Julia Dudar, and Christof Schöch (2022). "Evaluation of Measures of Distinctiveness. Classification of Literary Texts on the Basis of Distinctive Words". In: *Journal of Computational Literary Studies* 1 (1). [10.48694/jcls.102](https://doi.org/10.48694/jcls.102).

- Dunning, Ted (1994). "Accurate Methods for the Statistics of Surprise and Coincidence". In: *Computational Linguistics* 19 (1), 61–74.
- Fialho, Olivia (2019). "What Is Literature for? The Role of Transformative Reading". In: *Cogent Arts & Humanities* 6 (1). Ed. by Anezka Kuzmicova. [10.1080/23311983.2019.1692532](https://doi.org/10.1080/23311983.2019.1692532).
- Gabrielatos, Costas (2018). "Keyness Analysis". In: *Corpus Approaches to Discourse: A Critical Review*. Routledge, 225–258.
- Gitelman, Lisa (2013). '*Raw Data*' Is an Oxymoron. MIT Press. [10.7551/mitpress/9302.001.0001](https://doi.org/10.7551/mitpress/9302.001.0001).
- Gries, Stefan Th. (2008). "Dispersions and Adjusted Frequencies in Corpora". In: *International Journal of Corpus Linguistics* 13 (4), 403–437. [https://www.stgries.info/research/2008\\_STG\\_Dispersion\\_IJCL.pdf](https://www.stgries.info/research/2008_STG_Dispersion_IJCL.pdf) (visited on 10/07/2024).
- Hickman, Miranda B. (2012). "Introduction: Rereading the New Criticism". In: *Rereading the New Criticism*. Ed. by Miranda B. Hickman and John D. McIntyre. Ohio State University Press, 1–21. <https://core.ac.uk/download/pdf/159569564.pdf> (visited on 10/07/2024).
- Koolen, Marijn, Peter Boot, and Joris van Zundert (2020). "Online Book Reviews and the Computational Modelling of Reading Impact". In: *Proceedings of the Workshop on Computational Humanities Research*, 149–169. <http://ceur-ws.org/Vol-2723/long13.pdf> (visited on 10/03/2024).
- Koolen, Marijn, Olivia Fialho, Julia Neugarten, Joris van Zundert, Willem van Hage, Ole Mussmann, and Peter Boot (2023). "How Can Online Book Reviews Validate Empirical In-depth Fiction Reading Typologies?" In: *IGEL 2023 : Rhythm, Speed, Path: Spatiotemporal Experiences in Narrative, Poetry, and Drama*. <https://discourse.igelsociety.org/t/how-can-online-book-reviews-validate-empirical-in-depth-fiction-reading-typologies/370> (visited on 01/16/2024).
- Laurino Dos Santos, Henrique and Jonah Berger (2022). "The Speed of Stories: Semantic Progression and Narrative Success". In: *Journal of Experimental Psychology. General* 151 (8), 1833–1842. [10.1037/xge0001171](https://doi.org/10.1037/xge0001171).
- Lijffijt, Jeffrey, Terttu Nevalainen, Tanja Säily, Panagiotis Papapetrou, Kai Puolämäki, and Heikki Mannila (2016). "Significance Testing of Word Frequencies in Corpora". In: *Digital Scholarship in the Humanities* 31 (2), 374–397. [10.1093/lhc/fqu064](https://doi.org/10.1093/lhc/fqu064).
- Loi, Christina, Frank Hakemulder, Moniek Kuijpers, and Gerhard Lauer (2023). "On How Fiction Impacts the Self-concept: Transformative Reading Experiences and Storyworld Possible Selves". In: *Scientific Study of Literature* 12 (1), 44–67. [10.61645/ssol.181](https://doi.org/10.61645/ssol.181).
- Manovich, Lev (2013). *Software Takes Command*. International Texts in Critical Media Aesthetics. Bloomsbury Academic.
- Miall, David S. and Don Kuiken (1994). "Beyond Text Theory: Understanding Literary Response". In: *Discourse Processes* 17 (3), 337–352. [10.1080/01638539409544873](https://doi.org/10.1080/01638539409544873).
- Moreira, Pascale, Yuri Bizzoni, Kristoffer Nielbo, Ida Marie Lassen, and Mads Thomsen (2023). "Modeling Readers' Appreciation of Literary Narratives Through Sentiment Arcs and Semantic Profiles". In: *Proceedings of the the 5th Workshop on Narrative Understanding*, 25–35. <https://aclanthology.org/2023.wnu-1.5> (visited on 01/22/2024).
- Nguyen, Minh Van, Viet Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen (2021). "Trankit: A Light-Weight Transformer-based Toolkit for Multilingual Natural Language Processing". In: *Proceedings of the 16th Conference of the European Chapter of the*

- Association for Computational Linguistics: System Demonstrations.* [10.18653/v1/2021.eacl-demos.10](https://doi.org/10.18653/v1/2021.eacl-demos.10).
- Paquot, Magali and Yves Bestgen (2009). "Distinctive Words in Academic Writing: A Comparison of Three Statistical Tests for Keyword Extraction". In: *Corpora: Pragmatics and Discourse*. Brill, 247–269. [10.1163/9789042029101\\_014](https://doi.org/10.1163/9789042029101_014).
- Pichler, Axel and Nils Reiter (2022). "From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities". In: *Journal of Cultural Analytics* 7 (4). [10.22148/001c.57195](https://doi.org/10.22148/001c.57195).
- Prescott, Andrew (2023). "Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities?" In: *Digital Humanities Quarterly* 17 (2). <https://www.digit.alhumanities.org/dhq/vol/17/2/000689/000689.html> (visited on 01/16/2024).
- Rawson, Katie and Trevor Muñoz (2016). *Against Cleaning*. Project Blog. <http://curatingmenus.org/articles/against-cleaning/> (visited on 09/30/2016).
- Regis, Pamela (2003). *A Natural History of the Romance Novel*. University of Pennsylvania Press.
- Schöch, Christof (2017). "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama". In: *Digital Humanities Quarterly* 11 (2). <https://www.digit.alhumanities.org/dhq/vol/11/2/000291/000291.html> (visited on 10/07/2024).
- Sobchuk, Oleg and Artjoms Šēla (2023). "Computational Thematics: Comparing Algorithms for Clustering the Genres of Literary Fiction". In: *arXiv*. [10.48550/arXiv.2305.11251](https://arxiv.org/abs/10.48550/arXiv.2305.11251).
- Thompson, Laure and David Mimno (2018). "Authorless Topic Models: Biassing Models Away from Known Structure". In: *Proceedings of the 27th International Conference on Computational Linguistics*, 3903–3914. <https://aclanthology.org/C18-1329> (visited on 10/03/2024).
- Toubia, Olivier, Jonah Berger, and Jehoshua Eliashberg (2021). "How Quantifying the Shape of Stories Predicts Their Success". In: *Proceedings of the National Academy of Sciences* 118 (26), 1–5. [10.1073/pnas.2011695118](https://doi.org/10.1073/pnas.2011695118).
- Uglanova, Inna and Evelyn Gius (2020). "The Order of Things. A Study on Topic Modelling of Literary Texts". In: *Proceedings of the Workshop on Computational Humanities Research*, 57–76. <https://ceur-ws.org/Vol-2723/long7.pdf> (visited on 10/03/2024).
- Warnock, John (1978). "A Theory of Discourse, by James L. Kinneavy. (Review)". In: *Style* 12 (1), 52–54. <https://www.jstor.org/stable/45109026> (visited on 01/16/2024).
- Wimsatt, William K. (1954). "The Intentional Fallacy". In: *The Verbal Icon: Studies in the Meaning of Poetry*. University Press of Kentucky, 3–20. [https://www.sas.upenn.edu/~cavitch/pdf-library/WimsattBeardsley\\_Intentional.pdf](https://www.sas.upenn.edu/~cavitch/pdf-library/WimsattBeardsley_Intentional.pdf) (visited on 10/07/2024).
- Zundert, Joris van, Marijn Koolen, and Karina van Dalen-Oskam (2018). "Predicting Prose that Sells: Issues of Open Data in a Case of Applied Machine Learning". In: *JADH 2018 'Leveraging Open Data': Proceedings of the 8th Conference of Japanese Association for Digital Humanities*, 175–177. [https://conf2018.jadh.org/files/Proceedings\\_JADH2018\\_rev0911.pdf](https://conf2018.jadh.org/files/Proceedings_JADH2018_rev0911.pdf) (visited on 11/07/2018).
- Zundert, Joris van, Marijn Koolen, Julia Neugarten, Peter Boot, Willem van Hage, and Ole Mussmann (2022). "What Do We Talk About When We Talk About Topic?" In:

*Proceedings of Computational Humanities Research*, 398–410. [https://ceur-ws.org/Vo  
l-3290/](https://ceur-ws.org/Vol-3290/) (visited on 11/22/2023).

NUR code	NUR label	Genre label
280	Children's Fiction general	Children's fiction
281	Children's fiction 4–6 years	Children's fiction
282	Children's fiction 7–9 years	Children's fiction
283	Children's fiction 10–12 years	Children's fiction
284	Children's fiction 13–15 years	Young adult
285	Children's fiction 15+	Young adult
300	Literary fiction general	Literary fiction
301	Literary fiction Dutch	Literary fiction
302	Literary fiction translated	Literary fiction
305	Literary thriller	Literary thriller
312	Pockets popular fiction	Literary fiction
313	Pockets suspense	Suspense
330	Suspense general	Suspense
331	Detective	Suspense
332	Thriller	Suspense
334	Fantasy	Fantasy fiction
339	True crime	Suspense
342	Historical novel (popular)	Historical fiction
343	Romance	Romance
344	Regional and family novel	Regional fiction

**Table 2:** The selected NUR codes of novels in our dataset of 18,885 novels and their mapping to genres.

## A. Mapping NUR Codes to Genre Labels

The complete mapping of NUR codes to genre labels is shown in [Table 2](#).

## B. Overlap between Themes in Terms of Shared Books

The topic modeling process assigns each book to a single topic, but because individual topics can be linked to multiple themes, their books are also linked to multiple themes. As a consequence, themes share books and reviews, and some pairs of themes may have larger overlap than others. This overlap between themes is shown for pairs of themes where for one theme at least 25% of the books for one theme are shared by the other theme.

## C. Correlations between Themes in Terms of Impact

The correlations between themes in terms of the percent difference (%Diff) per impact term for generic *Affect*, *Narrative*, and *Aesthetics* are shown in [Figure 13](#), [Figure 14](#), and [Figure 15](#), respectively.

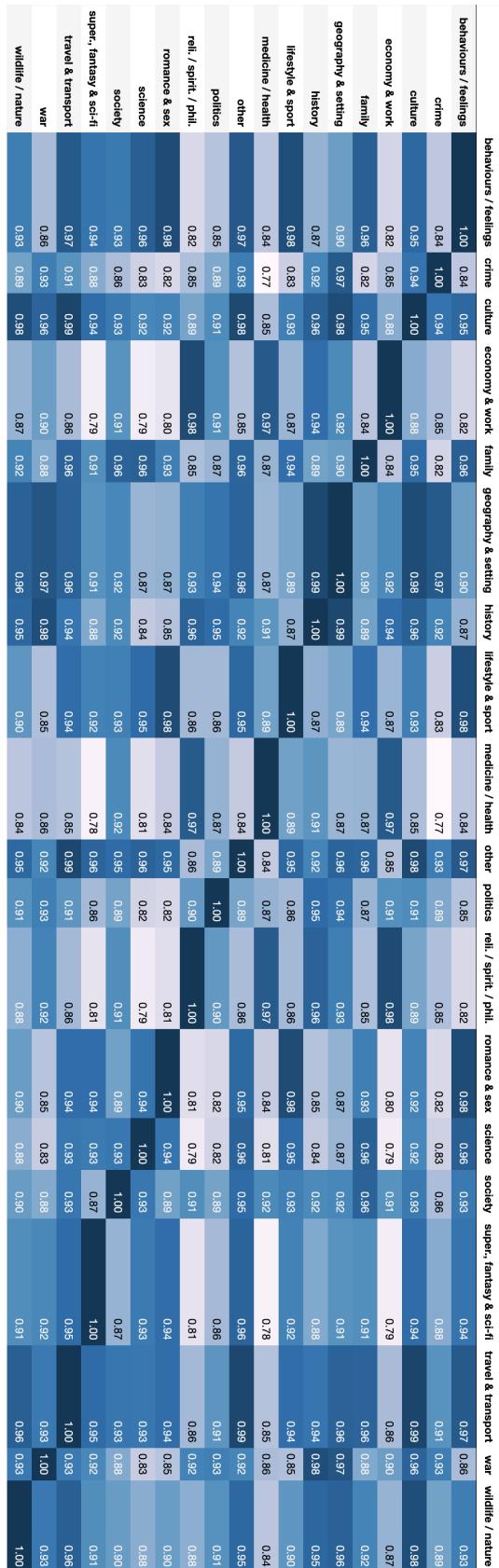
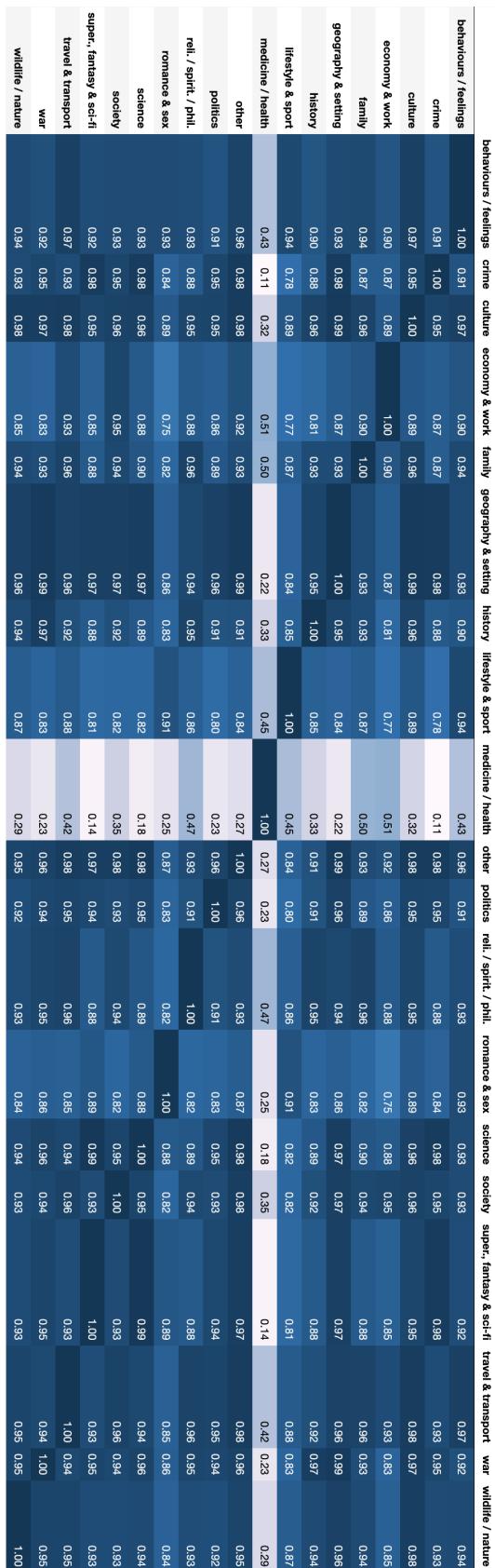
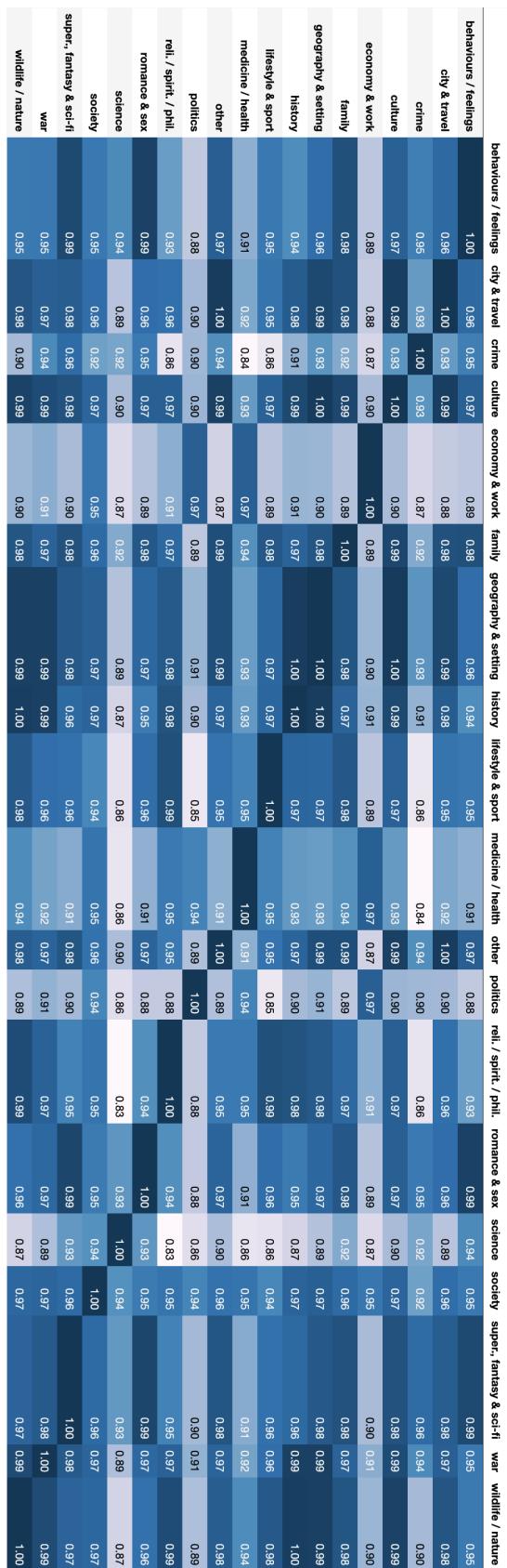


Figure 13: Percent different correlations between themes based on general Affect terms.

Figure 14: Percent different correlations between themes based on *Narrative impact terms*.



**Figure 15:** Percent different correlations between themes based on general Aesthetic impact terms.

Theme 1	Share 1	Theme 2	Share 2	Book Overlap	Books Theme 1	Books Theme 2
crime	0.33	geo. & setting	0.14	619	1,899	4,317
culture	0.49	geo. & setting	0.40	1,713	3,524	4,317
econ. & work	0.36	behav./feelings	0.12	446	1,232	3,860
econ. & work	0.30	society	0.44	371	1,232	851
econ. & work	0.25	politics	0.49	310	1,232	634
family	0.65	behav./feelings	0.08	324	498	3,860
family	0.30	culture	0.04	151	498	3,524
geo. & setting	0.40	culture	0.49	1,713	4,317	3,524
history	0.51	geo. & setting	0.24	1,038	2,020	4,317
history	0.31	war	0.65	622	2,020	952
life st. & sport	0.31	medi./health	0.20	216	702	1,058
politics	0.49	econ. & work	0.25	310	634	1,232
politics	0.49	society	0.36	310	634	851
society	0.44	econ. & work	0.30	371	851	1,232
society	0.36	politics	0.49	310	851	634
war	0.65	history	0.31	622	952	2,020

**Table 3:** Overlap in books between themes, for themes where one theme shares at least 25% of the books with the other theme.