

The event of autofiction in English: genre classification in an evolving literary system

Jeffrey Clapp¹ , Cheng Tsz Chung² , Ralph Fung¹ , and Lau Chaak Ming¹ 

¹ Education University of Hong Kong, Hong Kong SAR, China

² Kyushu University, Japan

Abstract

The genre system of English writing has recently been disrupted by the emergence of “autofiction”, generally understood as a hybrid of autobiographical and fictional narrative. Critics and commentators have sought to specify the nature of autofictional hybridization, and the set of texts to which the term might apply, yet without achieving much consensus. Because autofiction remains inchoate, and because it lies “between” fiction and nonfiction in ways that are as yet underspecified, we suggest that autofiction presents an interesting test case for commonly-employed genre classification methods in computational literary studies. We therefore contribute the first available quantitative, corpus-based examination of autofiction in English. Gathering 195 English-language works frequently described as autofiction(al) in various contexts, we explore the distinctiveness of these works relative to other kinds of contemporary fiction and nonfiction. Although with differing degrees of accuracy, we find that autofiction can be distinguished from fiction and nonfiction using most frequent words, LDA topics, and WordNet supersenses. We suggest that the word “autofiction” has been applied to a specific kind of contemporary writing—one in which *not much happens to a small number of people*. We further establish that autofiction can be identified via internal evidence, and not only via paratextual or extratextual factors.

Keywords: autofiction, contemporary literature, genre, genre classification, computational literary studies

1 Introduction

While the term “autofiction” has been in use in French since the 1970s, it has only recently been used in English [6; 10]. Since 2010, the term has become dramatically more common in English, occasioning an extended conversation about what it means, whether it is useful, and how it relates to other available concepts, like the “autobiographical novel” [23]. At the same time, regular use of the category has had the effect of gathering together a body of writing involving a large number of authors and works. The adoption of the term “autofiction” and the resulting evolution in the system of literary genres in English are the subject of this paper and the related project.

First French, and more recently Anglophone, literary critics and commentators have worked to specify the meaning of the word “autofiction”—without much consensus. Andersen has noted that for Philippe Gasparini, autofiction is a genre wherein the events are true and the fictive element is limited to “the very form of the narrative.” Whereas for Gérard Genette, autofiction is a genre in which “I, the author, am going to tell you a story of which I am the hero but which never happened to me” [2, p. 9]. The two definitions do not precisely invert one another. But they present substantially different accounts of what it might mean for autofiction to “hybridize” existing genres. For this reason, autofiction presents an interesting test case for genre classification methods employed in computational literary studies [1; 12; 19; 26; 27].

Jeffrey Clapp, Cheng Tsz Chung, Ralph Fung, and Lau Chaak Ming. “The event of autofiction in English: genre classification in an evolving literary system.” In: *Computational Cultural Science Workshop 2026*, ed. by Suzanne Mpouli, Jean-Baptiste Camps, Florian Cafiero, and Chahan Vidal-Gorène. Vol. 6. Anthology of Computers and the Humanities. 2026, 94–111. <https://doi.org/10.63744/td9AvQgRyaX7>.

© 2026 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

Alison Gibbons laments that “current criticism ultimately takes the form of figurative reasoning and unconfirmed theoretical speculation” [11, p. 474] and has sought to put the definition of autofiction on a firmer footing via the cognitive linguistics concept of “conceptual blending.” Authors write autofiction, and readers read it, by conjoining elements of a textual world with an empirical one. Authors provide points of connection—often, proper names—that knit these spaces together. Gibbons employs reader response data to suggest that readers really do discern “clues” that conjoin and distinguish textual and empirical worlds. Readers note that they are compelled to revise their understandings, often repeatedly, in the process of reading autofictional works.

The Novelist, a 2022 book by Jordan Castro, is among the well-attested recent publications in the corpus we analyze below. In this very brief novel, a novelist resembling the author spends a day at his laptop, trying to make progress on a book manuscript. He is repeatedly distracted by thoughts about another novelist, named Jordan Castro, whose work the novelist appreciates, but who has been sharply criticized online. In *The Novelist* it is not at all clear how and whether the diegetic “Jordan Castro,” or the first-person narrator, or both, correspond to the author whose name we find on the cover of the book.

If, as *The Novelist* suggests, autofiction is primarily about muddling readers’ sense of empirical and biographical veracity, it might seem impossible to grasp autofiction with the genre classification methods available to computational literary studies. Our research, however, was motivated by the suspicion that these methods, which ably parse fiction from nonfiction, or detective from mystery novels, among many other genres and subgenres, might well shed light on the emergence of autofiction as a genre formation in English. In thinking about literary genre, Ted Underwood has remarked that genre boundaries can be “blurry” and proposes that statistical methods offer means of understanding, rather than merely pondering, that blur [27, p. 19]. Underwood then gives as examples the fictionalized memoir and the novel *à clef*—two examples that might well be considered potential definitions of “autofiction.”

We hypothesize that many aspects of the works in our corpus, including thematic preoccupations, or tone, or style, may also motivate readers and critics to apply the word autofiction (or “autofictional”) to a work—even where narrators appear in the third person, or are extradiegetic, or do not share their names with authors. For example, as *The Novelist* suggests, when authors combine their lives and their work, stories of writing, publishing, and reception naturally come to the fore. However, as Hywel Dix has emphasized, this is by no means a universal element of autofiction. The story of “How the Author Became” can be less than central to a work which still confounds an author’s biography with fictional or fantastical material [9]. Considerations like these suggested to us that genre measures might well pick autofiction out from other kinds of writing, even if most definitions of autofiction seem to rely on extratextual information.

Given the difficulty of defining autofiction, more than a few have suggested that the concept is modish or superfluous. These claims motivate our approach. But we are also motivated by the ways that literary criticism has received the event of autofiction as evidence of decay in the imaginative capacity of contemporary literature. That pathology has primarily been linked to recent changes in political economy, understood generally as neoliberalization, or financialization, or platformization [13; 14; 16; 25; 28]. For example, McGurl remarks dismissively that “one cannot help but notice the strange kinship of minimalist autofiction with the contemporary mass-market romance novel” [16, pp. 219–20]. Wasserman laments that autofiction’s “commitment to interiority” can offer only the “latest literary preserve of the troubled self in troubled times” [28, p. 563]. Although this is not the place for a full discussion, it is fair to say that Anglophone literary criticism regards both autofictional texts and the “autofiction” moniker itself as evidence of something amiss in literature and its surroundings. Our project seeks to test some basic hypotheses: is there really a “strange kinship” between autofiction and the romance novel? Is autofiction more “interior” than other kinds of texts? Although the present paper cannot fully explore these areas, they motivate

our quantitative research into the event of autofiction in English. Our project seeks to model an emergent “literary system” [7]—insofar as this is possible with an inchoate and disputed genre category like this one.

We also offer some insight into the utility of genre classification methods in this particular context. We do not have and will not have ground-truth labels that identify literary texts as autofictions. Instead we observe a highly unstable set of ascriptions, described in more detail below. Daily the genre is both further consolidated and further diffused by statements that attach the names of authors and their works to the autofiction concept. By catching that process in action, we seek to add understanding to how genre functions in contemporary literary culture.

2 Corpus

We gather and prepare for analysis a corpus of 195 texts written in English, published since the year 2000, that have repeatedly been described as autofiction (or “autofictional”). The titles of these works are in Appendix I. In the present paper we refer to this corpus as AUTOFICTION or AUT. Our corpus includes those texts most frequently described as autofiction across four kinds of sources: 1) academic monographs, 2) academic journal article abstracts, 3) the online social reading platform Goodreads, and 4) on the open web, as scraped by Common Crawl. Ascriptions from five academic books about autofiction were collected by hand (two works, [29] and [10], include lists of autofictions). Around 800 journal article abstracts containing the keywords “autofiction” and “autofictional” were collected from Web of Science and EBSCO in April 2026. Goodreads tags were scraped from the “autofiction” and “autofictional” shelves in April 2026. Around 27,000 webpages containing the keywords were collected from FineWeb distillations of Common Crawl data in April 2026 [18]. An LLM-based extraction script was used to identify works specifically denoted as autofiction(al) by title in the abstracts and the web sources.

In order to arrive at a corpus which is both manageable and defensible, we include works that meet one or more of the following inclusion criteria: 1) described as autofiction(al) in three or more of the four source categories; 2) described as autofiction(al) in at least one of the academic and one of the online source categories; 3) described as autofiction(al) in two or more journal article abstracts; 4) described as autofiction(al) on four or more different webpages; 5) “shelved” as autofiction(al) four or more times on Goodreads. This set of criteria gave a list of 222 texts; we were able to acquire and pre-process 195 of them. One author (Rachel Cusk) has seven texts in the corpus; three have four texts each (Sheila Heti, Ben Lerner, Tao Lin); 129 authors have only one text. Although one might conjecture that different communities identify and discuss substantially different subsets of “autofiction,” this does not appear to be the case for the corpus discussed and analyzed below.

During the process of corpus assembly, we explored corpora of $n=116$ and $n=161$, constructed with different inclusion criteria. Both included texts with five or more Goodreads tags, three or more open web mentions, one or more academic ascriptions, or were described as autofiction(al) across two or more source categories. The $n = 116$ corpus was further limited to twenty-first century texts. Even though neither of these corpus slices were simply subsets of our current $n=195$ corpus, these did not give markedly different genre classification results. Still, it is important to understand that the construction of our “ascription corpus” shows extremely diffuse use of the term “autofiction.” The four source categories give more than 12,000 total generic ascriptions, together identifying more than 3,000 unique English-language texts as autofiction(al). Of these, only 300 are identified as autofiction(al) in more than one of the four source categories described above. This is already an interesting research result, one that gives insight into how this particular generic concept has been deployed.

Our current inclusion criteria result in a reasonably balanced corpus in terms of its authors’ demographic characteristics. These indications are sourced from Wikidata and may contain errors.

We do not attempt any analysis based on these metadata categories. Still, it may be useful to know that the 195 autofictions in this corpus are the work of 153 different authors, of whom 80 are female, 67 are male, and 6 identify otherwise. Eighty are primarily associated with the US, 31 with the UK, and 42 with other places.

We process our corpus via BookNLP [4], which tokenizes texts and then tags text features, including named entities and WordNet supersenses. Processing the corpus in this way enables comparability with the derived data published as the CONLIT corpus [20]. CONLIT presents data derived from book-length English-language texts published in the twenty-first century. Like our autofiction corpus, these texts are under copyright in multiple jurisdictions, so full-text datasets cannot be published or distributed. Each text in CONLIT is presented as belonging to a category (fiction or nonfiction) and to one of ten genres: bestsellers (BS); mysteries (MY); books reviewed in the New York Times (NYT); prizewinners (PW); romance (ROM); science fiction (SF); biography (BIO); history (HIST); memoirs (MEM); and an admixture of nonfiction, like self-help texts (MIX). After removal of duplicates, we work with a subset of CONLIT in which there are 1542 fictional and 819 nonfiction texts. CONLIT provides the background framework against which we can test the generic specificity of contemporary autofiction in English.

3 Characters and Events

To begin, we present an interesting descriptive result. The BookNLP pipeline conducts NER and coreference resolution for long texts. Included are tags for persons and “realis events” (meaning events that happen, rather than events that are imagined or described). These tags allow a quick descriptive analysis of how much activity takes place, and how many persons are involved, in CONLIT fiction, CONLIT nonfiction, and our AUTOFICTION. Figure 1 shows the results for characters and Figure 2 for events. Figures 1 and 2 also break down these categories by CONLIT genre, as well as for AUTOFICTION, described here as A21. Normalized character and event counts readily distinguish fiction from nonfiction within the CONLIT corpus. Fiction has more events, but fewer characters, than nonfiction. Moreover, in terms of its character density, autofiction is more like fiction, whereas in terms of its event density, autofiction is more like nonfiction. Table 1 gives the statistical validation. We used nonparametric pairwise comparisons (Mann-Whitney, two-sided) on these normalized measures.

These are important early results for a genre which has often been regarded as a “hybrid” of fictional and nonfictional writing, even if the terms of this hybridity differ from those presented by narratologists. They tend to confirm, while also complicating, elements of the critical reception of autofiction. Figure 1 bears out Mark McGurl’s remark about autofiction and romance, in terms of the number of characters (they have the fewest of any genre), but Figure 2 confounds it in terms of the number of events—romance has the most action of any fiction genre, while autofiction has far less. In this variety of contemporary literature, we can discern that *not much happens to a small number of people*.

Comparison	Rank-biserial r	p
Character density, AUT vs FIC	-0.105	< .05
Character density, AUT vs NON	-0.617	< .001
Event density, AUT vs FIC	-0.738	< .001
Event density, AUT vs NON	0.171	< .001

Table 1: Character and event density statistics. The rank-biserial r values indicate a larger difference between AUT and NON for character density, and between AUT and FIC for event density.

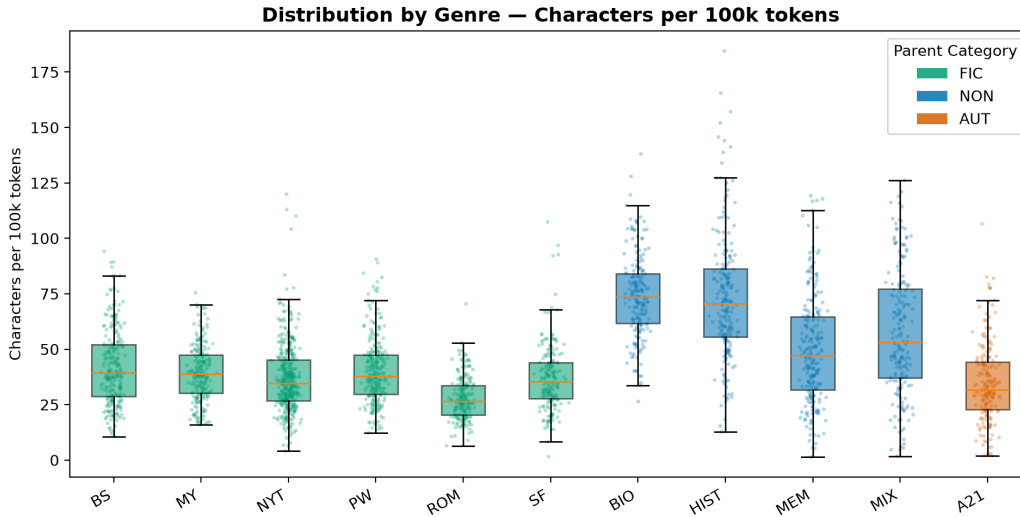


Figure 1: Character Density. Autofiction is more like fiction.

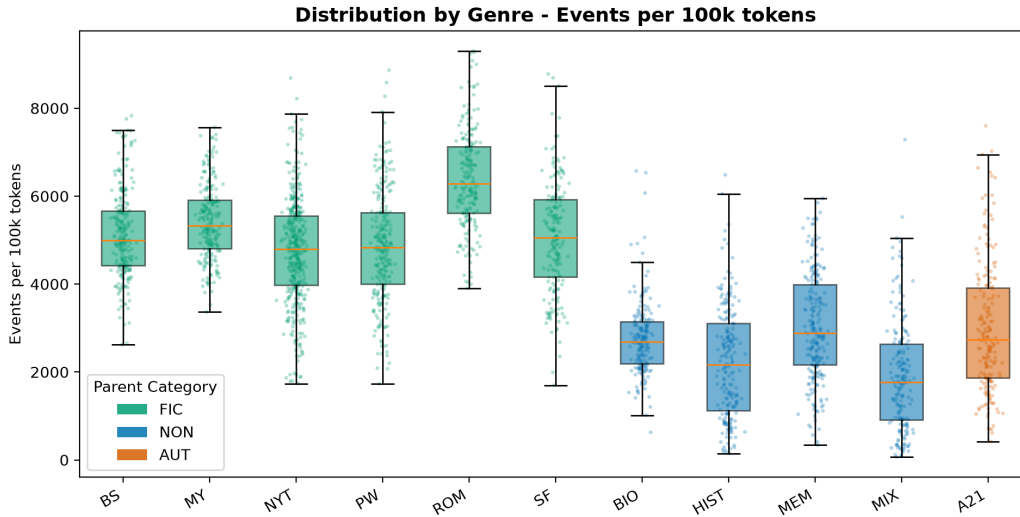


Figure 2: Event Density. Autofiction is more like non-fiction.

4 Classification Methods

In the following experiments we employ two parts of the CONLIT data: 1) the list of 10,000 most-frequent-words in the CONLIT corpus, with the per-book counts of those words; and 2) the per-book counts of words organized into the 41 top-level WordNet supersense categories. For example, these categories include “verb.communication” and “noun.time” [17]. Many function words have no supersense tag, so this method strips many frequently-used words, producing a topical or thematic abstraction of text content. The WordNet supersenses are tagged by BookNLP both in AUTOFICTION and in CONLIT, so these features are highly comparable. A similar procedure, instead using top-level categories from LIWC, has been shown to give interpretable and highly accurate results on genre classification tasks [19]. Supervised classification using WordNet features distinguishes every genre from every other genre with greater than 75% accuracy.

Classification using the CONLIT most-frequent-words list presents more challenges. CONLIT offers a list of the 10,000 MFW in the corpus, with per-book counts of each frequent word. Ideally, to conduct classification via MFW across CONLIT and AUTOFICTION, we would identify a new—

presumably slightly different-set of MFW from the combined fulltext corpus. However, because CONLIT fulltext is protected by copyright, this is not possible. We can only use MFW features by applying the CONLIT list to AUTOFICTION, which means that we explore AUTOFICTION in a CONLIT-defined feature space. This situation affects not only MFW features, but also the LDA topic modelling, which again cannot be performed on raw fulltext, but only on the per-MFW per-book counts provided by CONLIT. The only virtue of this situation is that, if the classification works at all, we can safely assume it would work better on fulltext data preprocessed in more directly comparable ways.

In order to proceed with classification using comparable features, we first lemmatize the CONLIT MFW list, then use spaCy to tag the CONLIT MFW list with parts-of-speech. The CONLIT derived data lacks context to distinguish POS very accurately, but we retain only the CONLIT MFW lemmas tagged as nouns, verbs, adjectives, and adverbs, and do not depend on the POS tags except as a filter. The result is a reduction of the MFW feature space to around 7700 words. In the case of MFW, although not in the case of topic modelling, we also slice out words that occur in fewer than five of the texts in the corpus, and those that occur in more than 95 percent of the texts in the corpus. These measures allow us to extract genre signal that remains after provenance signal resulting from preprocessing differences has been minimized.

Following these steps, we examine three feature types: Most Frequent Words; Topics, derived by Latent Dirichlet Allocation with the *tomotopy* library [5; 15]; and WordNet supersense counts. We use support vector machines with linear kernels. To prevent data leakage, all classifiers use 5-fold inner cross-validation for parameter optimization, then 10-fold outer cross-validation for actual classification. We report full-corpus results comparing FIC, NON, and AUT. We also ran balanced-subcorpus classification on the three feature sets and learned that the results do not vary significantly. We report the best-performing k value for classification by topics.

5 Classification Results

5.1 Pairwise Classification by SVM

Pairwise classification on these feature sets establishes that CONLIT fiction and nonfiction are highly distinguished from one another on these features, as Table 2 demonstrates. Slightly less accurate are the pairwise comparisons between autofiction and nonfiction. Noticeably less so are the comparisons between fiction and autofiction. MFW features are the best classifier of autofiction versus fiction and nonfiction, followed by strong results for supersenses, then weak results for topics. The latter may be related to the limited number of word features available, or it may be related to the open-ended thematic content of autofiction, which unlike other genres or subgenres is not linked to a specific iconography or setting. The MFW features likely preserve some provenance signal. This may explain why AUT is slightly more distinguishable from both CONLIT FIC and NON than they are from one another, using the MFW features.

Feature set	Setting	Accuracy	Macro-F1	Balanced accuracy
MFW POS lemmas	FIC vs NON	97.29	97.94	96.58
MFW POS lemmas	AUT vs NON	97.14	98.24	94.32
MFW POS lemmas	AUT vs FIC	97.41	98.55	90.03
Supersenses	FIC vs NON	95.17	96.30	94.67
Supersenses	AUT vs NON	87.96	74.58	89.42
Supersenses	AUT vs FIC	83.35	52.85	83.23
Topics (POS lemmas, k = 10)	FIC vs NON	70.00	57.44	70.00
Topics (POS lemmas, k = 10)	AUT vs NON	73.59	79.42	73.59
Topics (POS lemmas, k = 10)	AUT vs FIC	57.18	70.12	57.18

Table 2: Pairwise classifications.

5.2 Multiclass Classification by SVM

Multiclass SVM results place all three categories in the same feature space. Table 3 shows that classification results hold up reasonably well even on this more challenging task, with the exception of classification by topics. Even with the best-performing number of topics for this task (k = 50), these results show significant degradation.

Setting	Accuracy	Macro-F1	Balanced accuracy
MFW POS lemmas	95.34	91.73	89.56
Supersenses	80.09	71.71	80.29
Topics (POS lemmas, k = 50)	53.85	49.09	53.85

Table 3: Multiclass classification.

These classification results demonstrate clearly that autofiction can be identified using supervised classification methods, and on multiple kinds of text features. Given narratological accounts of autofiction that suggest that autofiction is defined by the empirical truth of its content, we regard this as an important result. The discovery that autofiction is more like fiction than nonfiction is an equally important indicator, so we pursue this below.

5.3 Projection experiments

In the following experiments we produce a CONLIT FIC-NON feature space and project AUTOFICTION into the space, in order to observe where it clusters relative to the SVM decision boundary. As the density plots in Figures 3, 4, and 5 show, different features give broadly similar results. In each case, AUTOFICTION forms a unimodal cluster near the decision boundary, with more texts leaning to the fiction side than to the nonfiction side. This lean is significantly stronger for the two kinds of features tied directly to corpus lemmas; the WordNet supersenses give a more symmetrical result. Table 4 shows where AUTOFICTION is distributed relative to the decision boundary in these FIC-NON spaces, as well as how broadly the three categories are spread in each case.

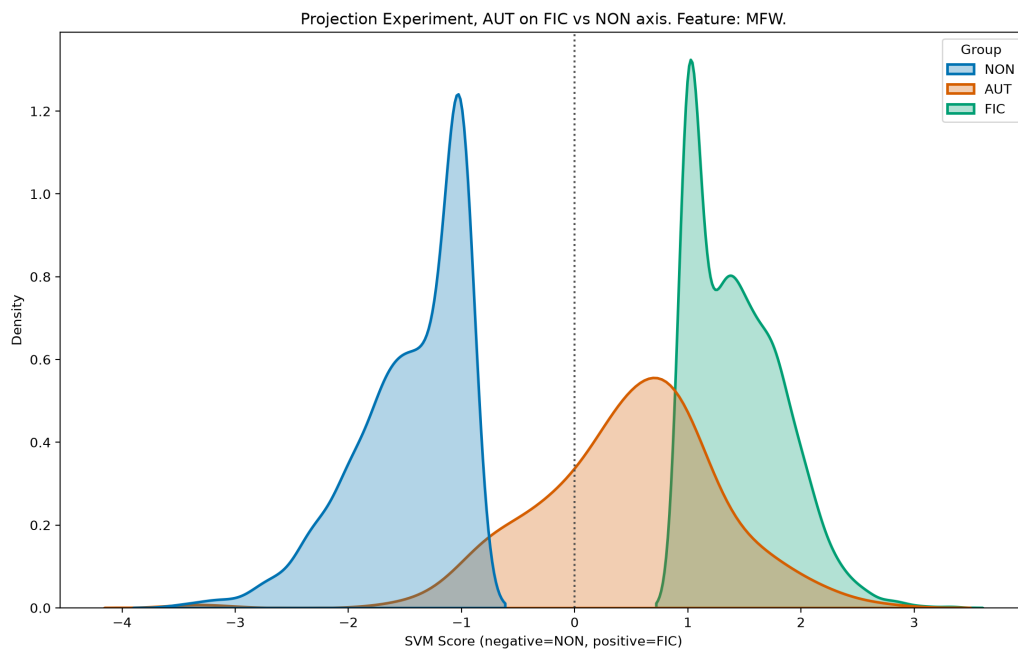


Figure 3: Projection with MFW features.

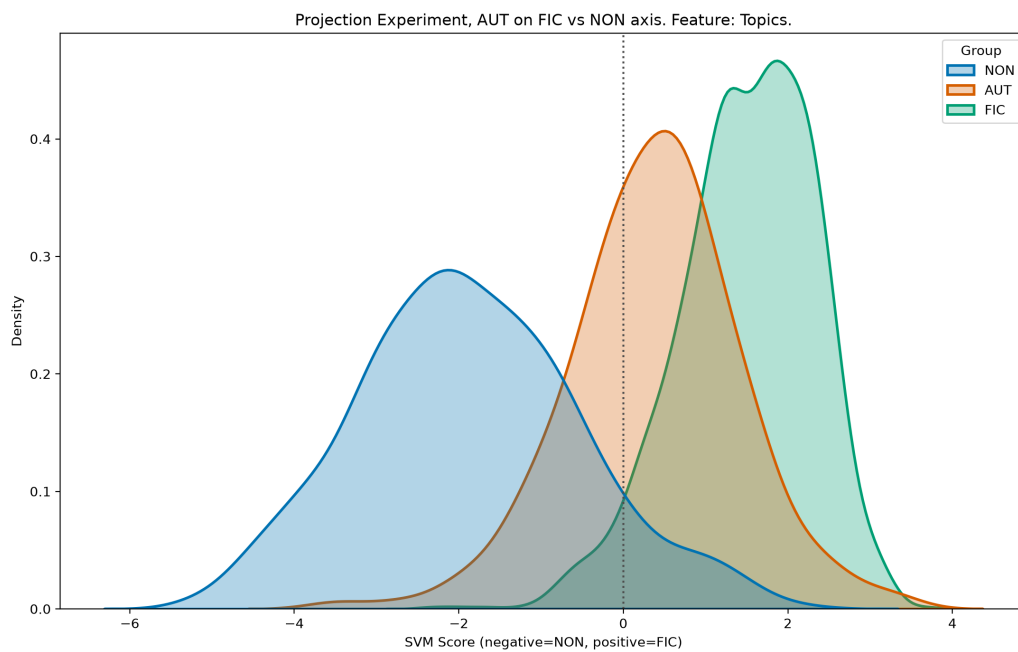


Figure 4: Projection with topic features.

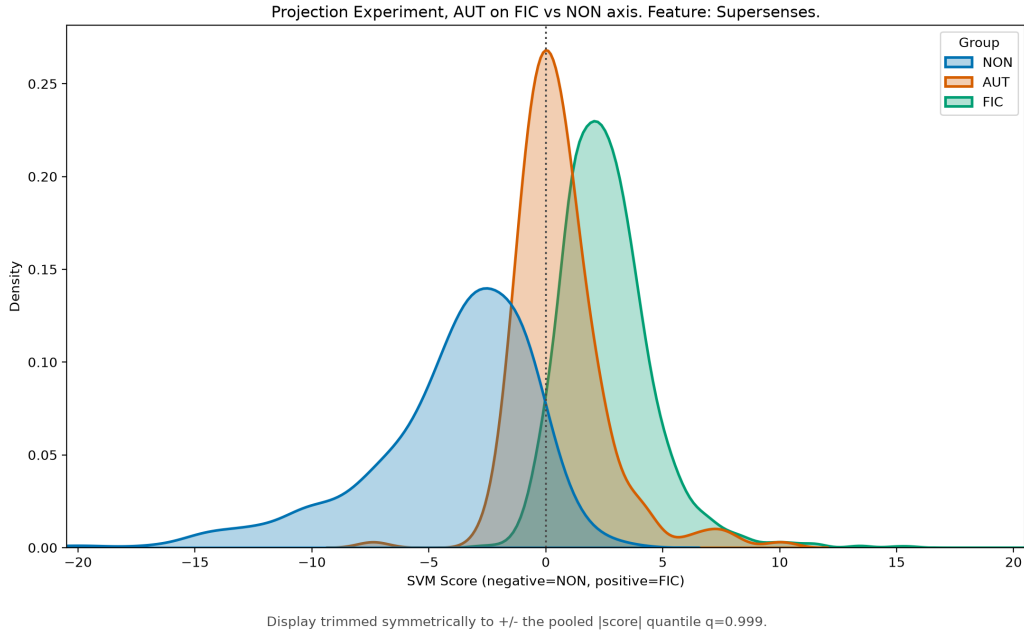
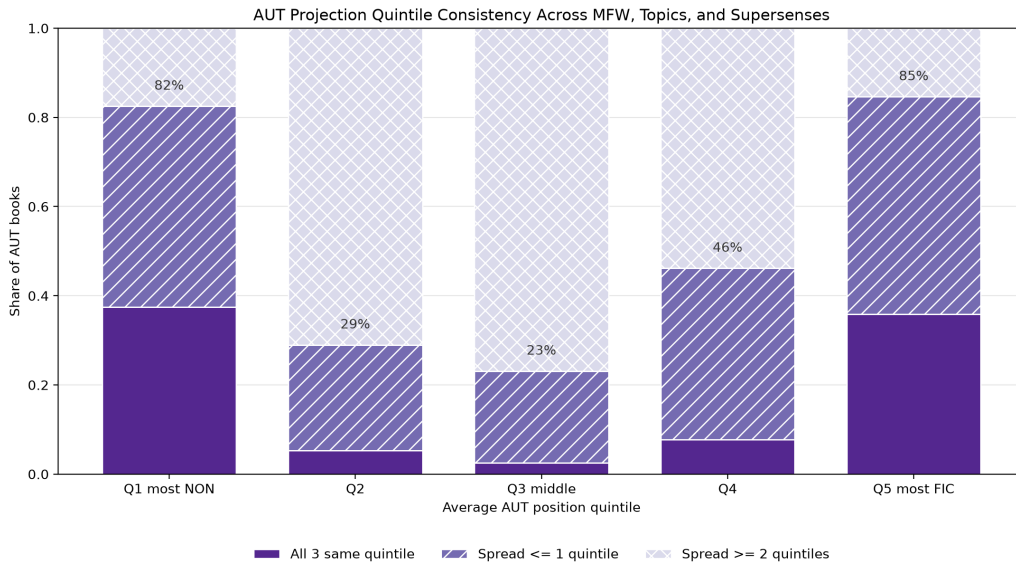


Figure 5: Projection with supersense features.

Feature set	AUT < 0	AUT > 0	AUT mean score	FIC SD	NON SD	AUT SD
MFW (POS lemmas)	47	148	0.484	0.405	0.490	0.797
Topics (POS lemmas, k = 10)	67	128	0.371	0.810	1.360	1.029
Supersenses	85	110	0.589	2.029	3.657	1.928

Table 4: Projection results.

These experiments order our 195 works of autofiction in a ranked list, from the negative-signed NON side to the positive-signed FIC side. Appendix I reports each text’s ranking in all three feature spaces. Comparing these three columns, we noticed significant commonalities in the relative placement of particular texts. *The Novelist* by Jordan Castro, for example, lands well on the FIC side of the boundary in all three cases. Figure 6 quantifies the fact that the set of works that appear in the tails of these distributions is quite stable. In other words, the most fiction- and nonfiction-like autofictions are consistent across these feature types, whereas the middle of the distribution is less stable.



Quintiles are computed separately within each feature set from signed AUT rank order; Q1 = most NON, Q5 = most FIC.

Figure 6: Consistency of projection results across features.

5.4 Unsupervised clustering

MFW, LDA topics, and WordNet supersenses do not effectively cluster these corpora into autofiction, fiction, and nonfiction. Across a variety of unsupervised clustering methods, clustering analyses produced differentiated two-cluster solutions with modest silhouette values. In the most interpretable clustering solutions, the data appeared to organise around a broad fiction/nonfiction tendency, with autofictional texts often aligning more closely with fiction rather than forming an independent cluster. These results contrast with other work in genre classification in literary studies, which shows reasonably interpretable outputs using similar features for hierarchical clustering for genre [24; 26]. Still, this result is not unexpected. We seek to understand an emergent evolution of the English-language literary field, and we do so on the basis of a highly diffuse, heterogeneous corpus of ascriptions.

6 Discussion

The genre system of English writing was recently disrupted by the emergence of “autofiction.” The present paper contributes first by understanding the event of autofiction as a discursive disruption that labelled many works in a new way, then by demonstrating that this event is reflected in some commonly-employed genre classification measures. The use of the term “autofiction” is not simply catachresis, nor does it simply denote a way of reading paratext, nor is it just a new means of selling books in a context defined by narcissism, or social media, or neoliberalism. Still, time will tell if this new species of expression will flourish in the evolving ecosystem of contemporary literature in English.

The SVM analysis indicates that fiction, nonfiction, and autofiction are distinguishable under supervised classification and via several different kinds of text features. This feature information does not produce three compact, self-organising clusters when the labels are withheld. Autofiction has become computationally legible, but obviously overlaps with both fiction and nonfiction and remains a hybrid or porous category. But this is not to say that we have made no progress. It is harder to distinguish autofiction from fiction than from nonfiction. In a feature space defined by

fiction and nonfiction, we find autofiction skewed toward fiction. So, to refer to the “blurry” genre kinds mentioned by Ted Underwood, it appears on the available evidence that autofiction is more like the “roman à clef” than the “fictionalized memoir” [27, p. 19]. Moreover, we have shown that we can make progress on claims like these via quantitative methods.

7 Limitations

In this contribution we have only sought a corpus sufficient to give statistical power and general coverage. We have not yet systematically investigated which texts, arising from which label sources, lead to the most robust classification results, nor have we leveraged any metadata, like author gender or year of publication. We have limited our analysis to texts professionally published in English since the year 2000, leaving aside not only self-publications, but also the very large number of texts published before 2000 to which the term “autofiction” has been retroactively applied. We have not employed all potential feature types that might be used for genre classification, nor have we used all potential variations of the frequent words, topics, and dictionary-based features we have employed. We are substantially limited by both our own ability to obtain texts, and by our limited access to the in-copyright materials that form CONLIT, our comparison corpus. Finally, we observe that we are working at the edges of the classification power of these feature sets. The meaningful classification and thus interpretation of *particular* texts, or authors, or imprints, is inhibited by all of these factors.

8 Conclusion and Future Work

Autofiction can be reliably distinguished from fiction and from nonfiction with some supervised methods. Projection experiments show autofiction lying “between” contemporary fiction and nonfiction with respect to the ways it uses words, themes, and concepts. Across text features, these experiments give a consistent account of which autofictions are more like fiction, and which are more like nonfiction.

In future work we will continue to examine the corpus in order to compare results to standing claims about autofiction in English. For example, we will systematically explore how autofiction relates to the other subgenres in the CONLIT corpus. It already appears that autofiction has a strong prestige signal that links it more closely to “prizewinning” books in the CONLIT corpus than to what Mark McGurl described as the “mass market romance novel.” Autofiction also bears less resemblance to “memoir” than might be imagined. We will also explore other features, including representations of power (as operationalized by “connotation frames” [3]) and affect (as operationalized by sentiment analysis). On the basis of recent improvements to character identification and coreference resolution, we will soon be able to address questions of community and individualism in autofictional texts [8]. We will consider how recent research into the difference between imagined and recalled narratives might relate to autofictional hybridity [21; 22]. And finally, we will consider ways in which LLM-based methods may make it possible to operationalize more directly the overlap between biographical and fictionalized content in contemporary autofiction in English.

Acknowledgements

Primary support for this project was provided by Hong Kong University Grants Council Grant # 18602924.

Data acquisition and preprocessing work was carried out on computer resources offered under the category of General Projects by the Research Institute for Information Technology, Kyushu University.

Research assistantship was provided by Emma Chan, Rosa Gurung, Cecelia Hu, Jing Lu, and Rai Ninam.

Appendix I Autofiction projection ranks

The table below gives the relative positions of each book in our current corpus in each projection experiment. Within each feature set, the 195 autofiction texts are ranked by their projection scores, from the most nonfiction-like (rank 1) to the most fiction-like (rank 195). Rows are displayed in MFW order.

<i>Title</i>	<i>Author</i>	MFW rank	Supersenses rank	Topics rank
<i>Dysphoria Mundi</i>	Paul B. Preciado	1	1	1
<i>The Lonely City</i>	Olivia Laing	2	2	20
<i>Just Kids</i>	Patti Smith	3	47	38
<i>The Outrun</i>	Amy Liptrot	4	22	155
<i>The Argonauts</i>	Maggie Nelson	5	37	10
<i>Torpor</i>	Chris Kraus	6	8	4
<i>H Is for Hawk</i>	Helen Macdonald	7	97	167
<i>Now That I See You</i>	Emma Batchelor	8	31	78
<i>I'm Glad My Mom Died</i>	Jennette McCurdy	9	72	52
<i>The Plot Against America</i>	Philip Roth	10	17	17
<i>A Minor Chorus</i>	Billy-Ray Belcourt	11	45	8
<i>What We Lose</i>	Zinzi Clemmons	12	52	40
<i>Mothertongues</i>	Ceridwen Dovey + Eliza Bell	13	7	95
<i>The Four Spent the Day Together</i>	Chris Kraus	14	5	9
<i>Diary of a Bad Year</i>	J. M. Coetzee	15	14	3
<i>Be Brief and Tell Them Everything</i>	Brad Listi	16	16	13
<i>Coventry: Essays</i>	Rachel Cusk	17	12	42
<i>Stay True</i>	Hua Hsu	18	11	12
<i>The Vulnerables</i>	Sigrid Nunez	19	50	6
<i>How to Write an Autobiographical Novel</i>	Alexander Chee	20	18	16
<i>Free: A Child and a Country at the End of History</i>	Lea Ypi	21	15	14
<i>Panpocalypse</i>	Carley Moore	22	71	69
<i>The Red Parts</i>	Maggie Nelson	23	81	61
<i>Everything I Know About Love</i>	Dolly Alderton	24	9	123
<i>The Absolutely True Diary of a Part-Time Indian</i>	Sherman Alexie	25	95	60
<i>The Year of Magical Thinking</i>	Joan Didion	26	13	5
<i>Sludge Utopia</i>	Catherine Fatima	27	6	11
<i>Only Americans Burn In Hell</i>	Jarett Kobek	28	3	2
<i>Crying in H Mart</i>	Michelle Zauner	29	59	138
<i>The Instant</i>	Amy Liptrot	30	30	96
<i>Why Be Happy When You Could Be Normal?</i>	Jeanette Winterson	31	34	108
<i>Either/Or</i>	Elif Batuman	32	153	27
<i>Leave Society</i>	Tao Lin	33	117	33
<i>In the Dream House</i>	Carmen Maria Machado	34	88	53
<i>A Life's Work: On Becoming a Mother</i>	Rachel Cusk	35	39	86
<i>Is There Still Sex in the City?</i>	Candace Bushnell	36	63	56
<i>100 Boyfriends</i>	Brontez Purnell	37	89	65
<i>Hotel</i>	Joanna Walsh	38	60	59
<i>In the End, It Was All About Love</i>	Musa Okwonga	39	41	91
<i>When I Hit You</i>	Meena Kandasamy	40	28	71
<i>Threshold</i>	Rob Doyle	41	19	97
<i>The Double Life of Liliane</i>	Lily Tuck	42	80	43
<i>Fuccboi</i>	Sean Thor Conroe	43	85	130
<i>Drifts</i>	Kate Zambreno	44	24	107
<i>Inferno</i>	Eileen Myles	45	79	46
<i>Real Estate</i>	Deborah Levy	46	53	157
<i>What Purpose Did I Serve in Your Life</i>	Marie Calloway	47	146	94
<i>Gone with the Mind</i>	Mark Leyner	48	56	29
<i>Fever</i>	Jonathan Bazzi	49	26	115
<i>America del Norte</i>	Nicolás Medina Mora	50	21	15
<i>Days of Distraction</i>	Alexandra Chang	51	113	32
<i>Cold Enough for Snow</i>	Jessica Au	52	122	126
<i>Monkey Boy</i>	Francisco Goldman	53	133	114
<i>Juliet the Maniac</i>	Juliet Escoria	54	160	127
<i>The Scent of Light</i>	Kristjana Gunnars	55	142	110
<i>What Does It Feel Like?</i>	Sophie Kinsella	56	69	134
<i>Small Rain</i>	Garth Greenwell	57	157	160
<i>Departure(s)</i>	Julian Barnes	58	23	21

Continued on next page

Table 5 (continued)

Title	Author	MFW rank	Supersenses rank	Topics rank
<i>Lying</i>	Lauren Slater	59	120	125
<i>Homeland Elegies</i>	Ayad Akhtar	60	112	22
<i>Memoirs and Misinformation</i>	Jim Carrey and Dana Vachon	61	126	77
<i>The Friend</i>	Sigrid Nunez	62	38	83
<i>M Train</i>	Patti Smith	63	96	92
<i>Please Miss</i>	Grace Lavery	64	67	45
<i>What Are You Going Through</i>	Sigrid Nunez	65	68	50
<i>A Heartbreaking Work of Staggering Genius</i>	Dave Eggers	66	154	31
<i>I Love You But I've Chosen Darkness</i>	Claire Vaye Watkins	67	107	103
<i>The Cost of Living: A Working Autobiography</i>	Deborah Levy	68	57	162
<i>Tremor</i>	Teju Cole	69	20	7
<i>Youth</i>	J. M. Coetzee	70	25	89
<i>Anam</i>	André Dao	71	43	51
<i>Inside Story</i>	Martin Amis	72	33	37
<i>Immigrant, Montana</i>	Amitava Kumar	73	40	18
<i>jPod</i>	Douglas Coupland	74	44	93
<i>Reproduction</i>	Louisa Hall	75	76	81
<i>Educated</i>	Tara Westover	76	173	147
<i>Book of Numbers</i>	Joshua Cohen	77	10	87
<i>Mean</i>	Myriam Gurba	78	130	116
<i>Alphabetical Diaries</i>	Sheila Heti	79	29	104
<i>Motherhood</i>	Sheila Heti	80	32	64
<i>Fake Accounts</i>	Lauren Oyler	81	98	49
<i>Bleaker House</i>	Nell Stevens	82	111	154
<i>A Particular Kind of Black Man</i>	Tope Folarin	83	149	67
<i>No One Is Talking About This</i>	Patricia Lockwood	84	105	119
<i>Mercy Gene</i>	JD Derbyshire	85	54	76
<i>10:04</i>	Ben Lerner	86	101	25
<i>Who They Was</i>	Gabriel Krauze	87	176	191
<i>The Idiot</i>	Elif Batuman	88	188	88
<i>Split Tooth</i>	Tanya Tagaq	89	102	179
<i>Mrs Gaskell & Me</i>	Nell Stevens	90	116	132
<i>Parade</i>	Rachel Cusk	91	84	48
<i>Outline of My Lover</i>	Douglas Martin	92	82	41
<i>Kudos</i>	Rachel Cusk	93	108	70
<i>Sight</i>	Jessie Greengrass	94	27	62
<i>Broken People</i>	Sam Lansky	95	156	149
<i>Cherry</i>	Nico Walker	96	147	141
<i>Alfred and Emily</i>	Doris Lessing	97	110	145
<i>Want</i>	Lynn Steger Strong	98	152	99
<i>My Lesbian Novel</i>	Renee Gladman	99	51	63
<i>The Pale King</i>	David Foster Wallace	100	158	58
<i>A Previous Life</i>	Edmund White	101	139	113
<i>Taipei</i>	Tao Lin	102	159	44
<i>A Lie Someone Told You about Yourself</i>	Peter Ho Davies	103	77	106
<i>All Fours</i>	Miranda July	104	136	153
<i>LIVEBLOG</i>	Megan Boyle	105	194	142
<i>The Topeka School</i>	Ben Lerner	106	118	26
<i>Weather</i>	Jenny Offill	107	73	72
<i>The Man Who Never Stopped Sleeping</i>	Aharon Appelfeld	108	83	55
<i>A Million Little Pieces</i>	James Frey	109	185	101
<i>Afternoon Raag</i>	Amit Chaudhuri	110	48	139
<i>The Eulogy</i>	Jackie Bailey	111	123	180
<i>Death Valley</i>	Melissa Broder	112	115	169
<i>The Suicide Museum</i>	Ariel Dorfman	113	4	23
<i>At Certain Points We Touch</i>	Lauren John Joseph	114	137	146
<i>Everything Is Illuminated</i>	Jonathan Safran Foer	115	178	137
<i>The Hero of This Book</i>	Elizabeth McCracken	116	91	82
<i>Leaving the Atocha Station</i>	Ben Lerner	117	129	34
<i>How Should a Person Be?</i>	Sheila Heti	118	92	102
<i>Richard Yates</i>	Tao Lin	119	150	80
<i>Old School Indian</i>	Aaron John Curtis	120	161	131
<i>Summertime</i>	J. M. Coetzee	121	42	75
<i>Things I Don't Want to Know</i>	Deborah Levy	122	93	185
<i>My Name Is Lucy Barton</i>	Elizabeth Strout	123	128	57
<i>Sanatorium</i>	Abi Palmer	124	62	176

Continued on next page

Table 5 (continued)

Title	Author	MFW rank	Supersenses rank	Topics rank
<i>Open City</i>	Teju Cole	125	66	35
<i>The Chronology of Water</i>	Lidia Yuknavitch	126	148	135
<i>Dept. of Speculation</i>	Jenny Offill	127	78	24
<i>The Novelist</i>	Jordan Castro	128	106	73
<i>Pure Colour</i>	Sheila Heti	129	49	84
<i>Lost Children Archive</i>	Valeria Luiselli	130	174	47
<i>Lion</i>	Sonya Walger	131	138	190
<i>Theory & Practice</i>	Michelle de Kretser	132	70	156
<i>The Brooklyn Follies</i>	Paul Auster	133	36	79
<i>Bluets</i>	Maggie Nelson	134	61	39
<i>Will There Ever Be Another You</i>	Patricia Lockwood	135	145	121
<i>A Line Made by Walking</i>	Sara Baume	136	144	182
<i>The Mobius Book</i>	Catherine Lacey	137	94	68
<i>Moonglow</i>	Michael Chabon	138	181	129
<i>A Lie About My Father</i>	John Burnside	139	99	111
<i>Freedom</i>	Jonathan Franzen	140	121	28
<i>Crudo</i>	Olivia Laing	141	65	109
<i>The Tragedy of Arthur</i>	Arthur Phillips	142	46	54
<i>In a Strange Room</i>	Damon Galgut	143	90	150
<i>Such a Fun Age</i>	Kiley Reid	144	170	120
<i>Black Wave</i>	Michelle Tea	145	134	143
<i>The Sarah Book</i>	Scott McClanahan	146	165	140
<i>A Manual for Cleaning Women</i>	Lucia Berlin	147	167	172
<i>Every Day Is for the Thief</i>	Teju Cole	148	55	30
<i>A Tale for the Time Being</i>	Ruth Ozeki	149	179	174
<i>I Am Not Sidney Poitier</i>	Percival Everett	150	184	19
<i>Preparation for the Next Life</i>	Atticus Lish	151	192	187
<i>If an Egyptian Cannot Speak English</i>	Noor Naga	152	127	173
<i>Forest Dark</i>	Nicole Krauss	153	103	164
<i>Grand Union</i>	Zadie Smith	154	119	74
<i>The Blazing World</i>	Siri Hustvedt	155	141	36
<i>Aftermath: On Marriage and Separation</i>	Rachel Cusk	156	86	122
<i>Lunar Park</i>	Bret Easton Ellis	157	171	118
<i>Transit</i>	Rachel Cusk	158	135	133
<i>The Wallcreeper</i>	Nell Zink	159	74	66
<i>Sea of Tranquility</i>	Emily St. John Mandel	160	104	90
<i>Saltwater</i>	Jessica Andrews	161	162	193
<i>Stubborn Archivist</i>	Yara Rodrigues Fowler	162	125	186
<i>Outline</i>	Rachel Cusk	163	131	98
<i>Walking to Hollywood</i>	Will Self	164	164	158
<i>Brat</i>	Gabriel Smith	165	172	105
<i>Checkout 19</i>	Claire-Louise Bennett	166	143	168
<i>Assembly</i>	Natasha Brown	167	87	124
<i>Shantaram</i>	Gregory David Roberts	168	195	183
<i>Friend of My Youth</i>	Amit Chaudhuri	169	58	100
<i>Break.up</i>	Joanna Walsh	170	35	151
<i>Memories of the Future</i>	Siri Hustvedt	171	177	136
<i>The Pregnant Widow</i>	Martin Amis	172	169	161
<i>Freshwater</i>	Akwaeke Emezi	173	155	177
<i>Notable American Women</i>	Ben Marcus	174	64	128
<i>Shoplifting from American Apparel</i>	Tao Lin	175	114	117
<i>The Small Backs of Children</i>	Lidia Yuknavitch	176	124	166
<i>A Ghost in the Throat</i>	Doireann Ní Ghríofa	177	109	175
<i>If You Love It, Let It Kill You</i>	Hannah Pittard	178	132	144
<i>On Earth We're Briefly Gorgeous</i>	Ocean Vuong	179	175	184
<i>The Patrick Melrose Novels</i>	Edward St. Aubyn	180	183	148
<i>Americanah</i>	Chimamanda Ngozi Adichie	181	182	159
<i>That Reminds Me</i>	Derek Owusu	182	100	170
<i>Transcription</i>	Ben Lerner	183	75	85
<i>7½</i>	Christos Tsiolkas	184	166	181
<i>Borges and Me</i>	Jay Parini	185	140	163
<i>Cleanness</i>	Garth Greenwell	186	168	178
<i>The Shards</i>	Bret Easton Ellis	187	190	171
<i>Beautiful World, Where Are You</i>	Sally Rooney	188	186	112
<i>What Belongs to You</i>	Garth Greenwell	189	163	152
<i>We the Animals</i>	Justin Torres	190	151	192

Continued on next page

Table 5 (continued)

<i>Title</i>	Author	MFW rank	Supersenses rank	Topics rank
<i>Bunny</i>	Mona Awad	191	187	189
<i>The Boat</i>	Nam Le	192	189	188
<i>Shuggie Bain</i>	Douglas Stuart	193	193	194
<i>Young Mungo</i>	Douglas Stuart	194	191	195
<i>Real Life</i>	Brandon Taylor	195	180	165

References

- [1] Allison, Sarah, Heuser, Ryan, Jockers, Matthew, Moretti, Franco, and Witmore, Michael. “Quantitative Formalism: An Experiment”. In: *Stanford Literary Lab Pamphlet*, no. 1 (2011).
- [2] Andersen, Claus Elholm. *Knausgaard and the Autofictional Novel*. State University of New York Press, 2023.
- [3] Antoniak, Maria, Field, Anjalie, Mun, Jimin, Walsh, Melanie, Klein, Lauren, and Sap, Maarten. “Riveter: Measuring Power and Social Dynamics Between Entities”. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 377–388. DOI: 10.18653/v1/2023.acl-demo.36.
- [4] Bamman, David. “BookNLP: A Natural Language Processing Pipeline for Books”. Version 1.0. 2020. URL: <https://github.com/booknlp/booknlp>.
- [5] Blei, David M, Ng, Andrew Y, and Jordan, Michael I. “Latent dirichlet allocation”. In: *Journal of machine learning research* 3 (2003), pp. 993–1022.
- [6] Bloom, Myra. “Sources of the Self(Ie): An Introduction to the Study of Autofiction in English”. In: *ESC: English Studies in Canada* 45, no. 1–2 (2019), pp. 1–18. DOI: 10.1353/esc.2019.0000.
- [7] Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press, 2018.
- [8] Bourgois, Antoine, Barré, Jean, Seminck, Olga, and Poibeu, Thierry. “Toward an Ontological Representation of Fictional Characters”. In: *Computational Humanities Research* 2 (2026). DOI: 10.1017/chr.2026.10025.
- [9] Dix, Hywel. *Autofiction and Cultural Memory*. London: Routledge, 2022. DOI: 10.4324/9781003313465.
- [10] Hywel Dix, edited by. *Autofiction in English*. Springer, 2018.
- [11] Gibbons, Alison. “A Cognitive Model of Reading Autofiction”. In: *English Studies* 103, no. 3 (Apr. 3, 2022), pp. 471–493. DOI: 10.1080/0013838X.2022.2050611.
- [12] Hatzel, Hans Ole, Stierner, Haimo, Biemann, Chris, and Gius, Evelyn. “Machine Learning in Computational Literary Studies”. In: *it - Information Technology* 65, no. 4-5 (2023), pp. 200–217.
- [13] Konstantinou, Lee. “Autofiction and Autoreification”. Feb. 6, 2021. URL: <https://www.leekonstantinou.com/2021/02/06/autofiction-and-autoreification/>.
- [14] Kornbluh, Anna. *Immediacy, or The Style of Too-Late Capitalism*. Verso, 2024.
- [15] Lee, Minchul. “bab2min/tomotopy: 0.12.3”. Version v0.12.3. 2022. DOI: 10.5281/zenodo.6868418.
- [16] McGurl, Mark. *Everything and Less: The Novel in the Age of Amazon*. Verso, 2021.
- [17] Miller, George A. “WordNet: a lexical database for English”. In: *Communications of the ACM* 38, no. 11 (1995), pp. 39–41.
- [18] Penedo, Guilherme, Kydlíček, Hynek, allal, Loubna Ben, Lozhkov, Anton, Mitchell, Margaret, Raffel, Colin, Werra, Leandro Von, and Wolf, Thomas. “The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale”. Oct. 31, 2024. URL: <http://arxiv.org/abs/2406.17557>.

- [19] Piper, Andrew. *Enumerations: Data and Literary Study*. University of Chicago Press, 2018.
- [20] Piper, Andrew. “The CONLIT Dataset of Contemporary Literature”. In: *Journal of Open Humanities Data* 8 (2022), pp. 1–24.
- [21] Piper, Andrew, Hamilton, Sil, Zhou, Haiqi, and Pianzola, Federico. “Fiction Flows: A Replication and Reinterpretation of Narrative Sequentiality”. In: *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, July 2026, pp. 34158–34174. DOI: 10.18653/v1/2026.acl-long.1576.
- [22] Sap, Maarten, Jafarpour, Anna, Choi, Yejin, Smith, Noah A., Pennebaker, James W., and Horvitz, Eric. “Quantifying the Narrative Flow of Imagined versus Autobiographical Stories”. In: *Proceedings of the National Academy of Sciences* 119, no. 45 (Nov. 8, 2022). DOI: 10.1073/pnas.2211715119.
- [23] Saunders, Max. “Autofiction, Autobiografiction, Autofabrication, and Heteronymy: Differentiating Versions of the Autobiographical”. In: *Biography* 43, no. 4 (2020), pp. 763–780.
- [24] Schöch, Christof. “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama”. In: *Digital Humanities Quarterly* (Mar. 24, 2017). DOI: 10.48550/arXiv.2103.13019.
- [25] Sinykin, Dan. *Big Fiction: How Conglomeration Changed the Publishing Industry and American Literature*. New York: Columbia University Press, 2023.
- [26] Sobchuk, Oleg and ŠeĽa, Artjoms. “Computational Thematics: Comparing Algorithms for Clustering the Genres of Literary Fiction”. In: *Humanities and Social Sciences Communications* 11, no. 1 (Mar. 20, 2024). DOI: 10.1057/s41599-024-02933-6.
- [27] Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.
- [28] Wasserman, Sarah. “Critical Darlings, Critical Dogs: Joseph O’Neill and What Contemporary Criticism Doesn’t Want”. In: *American Literary History* 34, no. 2 (May 18, 2022), pp. 561–585. DOI: 10.1093/alh/ajac066.
- [29] Worthington, Marjorie. *The Story of “Me”: Contemporary American Autofiction*. University of Nebraska Press, 2018.