How Long was the Three Volume Novel?

Chris Forster

2023-08-23

I've been working with data related to the three-volume novel recently. I'll be writing a few posts as I try to tie this project together. In this post I'd like to talk just about *the length* of three-volume novels, and a bit of a challenge I encountered.

The Three-Volume Novel

The three-volume novel (or "triple-decker") is a distinctly Victorian format for the publication of fiction. While multivolume novels are as old as the form itself (*Tristram Shandy* [1759-1767] was initially in nine volumes; Richardson's *Clarissa* [1748] was published in seven volumes), the three-volume format came to be dominant in the Victorian period.

Philip Gaskell summarizes the three-volume novel like this:

> The mid-nineteenth-century English three-decker was typically a cloth-cased post octavo with about 20 gatherings, 320 pages, per volume. The edges were often uncut, and the type was small pica, or less commonly pica, with up to a nonpareil of lead

between the lines (i.e. 10 or 12 pt., up to 6pt. leaded). There were 900-920 pages of text, containing 150,000-200,000 words, padded out with heavy leading, wide margins, and extravagant chapter divisions. The median length of a large sample of three-deckers was actually 168,000 words, dividied into 45 chapters... Not all mid-nineteenth-century fiction was first published in three-decker form. About a third of the novels that were first published in book form came out in one or two (or very occasionally four) volumes. (301)[1]

This dominance was a function of private lending libraries. Novels in three volumes were quite expensive (one and a half guineas, or one pound eleven shillings and six pence).The "guinea" is an odd measure of value, equal to a pound a shilling. It was, per the OED, "the ordinary unit for a professional fee and for a subscription to a society or institution" ("Guinea"). That is—expressing the price in guineas has class connotations, which are not irrelevant here. Because they were so expensive, readers did not purchase triple-deckers—they borrowed them, one volume at a time typically, from private lending libraries like Mudie's. The three-volume novel became the standard format in which such libraries purchased novels.[2]. The three-volume novel, however, had disappeared by the turn of the twentieth century.

The three-volume novel therefore represents a rigid aesthetic and financial structure. It integrates a method of circulation and "monetization" (the subscription library) with an aesthetic/narrative mode.

---

[1] I need to check Gaskell—this quotation comes from a note in my Zotero library; I assume he is citing Lauterbach for this data, but not sure.

[2] Such libraries were crucial instituations of Victorian middle-class reading. If a library opted not to purchase a novel (as famously happened to George Moore), they could drmatically undermine its potential. This power of the the libraries outlived the three-volume novel format. As Nicola Wilson shows, such libraries operated as gatekeepers (and, essentially, censors), well into the first decade of the twentieth century (Wilson)

To crudely adapt some Marxist vocabulary, the three-volume novel is an interesting test case for broad questions about the relation of narrative superstructure to its enabling market base. Just as today we might ask "why does everything on Netflix look like that?" (Jackson), we might wonder if it is possible to detect a homogeneizing or otherwise shaping effect of the circulating library on the narratives and aesthetics of the British novel in the Victorian.

I am exploring such questions by taking advantage of Troy Bassett's invaluable bibliography *At the Circulating Library* for a list of titles, and culling data from the volumes available via HathiTrust . Searching through HathiTrust (using the aggregated HathiFiles to find titles listed in *At the Circulating Library* has been the most technically complicated part of this process. That, however, is a separate issue that I will skip over now. Let me simply say that I was able to locate 956 complete three-volume novels (represented by 2868 individual volumes), as well as 105 single volume novels from the *ATCL* bibliography, in HathiTrust.

We can use these titles to begin to investigate some of these questions. In this post I'd look just at a single, apparently simple, and ruthlessly quantitative, measure of the three-volume. How long were they?

Lengths

The *ATCL* bibliography contains more than 5,000 three-volume novels. I have located (what I believe are) 956 of those titles in HathiTrust and counted their lengths using the following basic code. Using the Python `htrc_features` library, one can count the words in a volue with this code.

```
tokens = vol.tokens_per_page(section='body').sum()
```

Once I had found about a thousand novels and computed their lengths I started graphing 'em. (I'll share those graphs in a moment.) But I wanted some check to see if my counts seemd

right. Thankfully, mine is not the first attempt to quantify the length of the three-volume novel. In 1957, Charles and Edward Lauterbach published an analysis of the three-volume novel in *The Papers of the Bibliographical Society of America*. They attempt to quantify the three-volume novel in terms of length (in words and pages) as well as more fine-grained measures such as the size of the page, size of the type area, typeface size, and so on. Their data, summarizing these elements for 105 novels, is impressive. How, you may wonder, did they manage this in a predigital age? They explain,

> Making the word count of a hundred five nineteenth-century would be impossible without mechanical aids and short-cuts. An electrical counter was developed which made it possible to count about 10,000 words an hour with only slight error. (263)

An electrical counter? Exactly what this means is not clear. In a footnote they promise "A paper describing this eletrical bibliographic aid," but I have found no record of this paper beyond this footnote—and so, for now, I remain somewhat uncertain in exactly *how* the Lauterbachs counted all these words.

However, the Lauterbach data provide a useful point of comparison. I used a trial of Amazon's OCR service to convert the tables from Lauterbach's publication, cleaned them up a bit (CSV). The mean length of a novel in Lauterbach's collection is 171,038 words (median, 170,493 words).You can get larger versions of the plots on this page by right-clicking and opening the image in a new tab.

The data I derived from HathiTrust looks like this, with a mean of 194,818 words (median 190,427 words).

Because of the very different number of titles (105 in Lauterbauch, and almost 1000 in the HathiTrust-derived dataset), a boxplot makes a comparison clearer.

I had expected (and indeed hoped) that the Lauterbach data
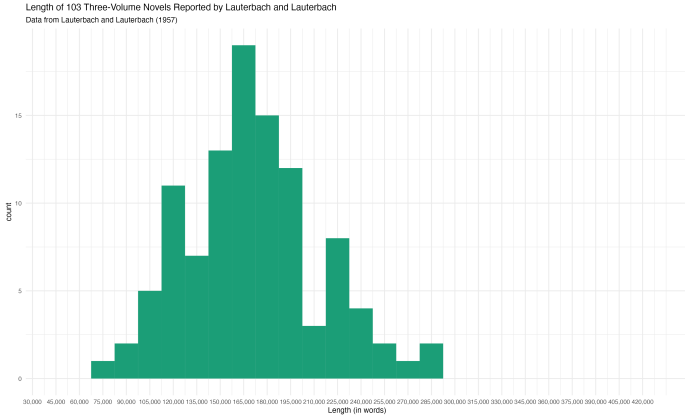
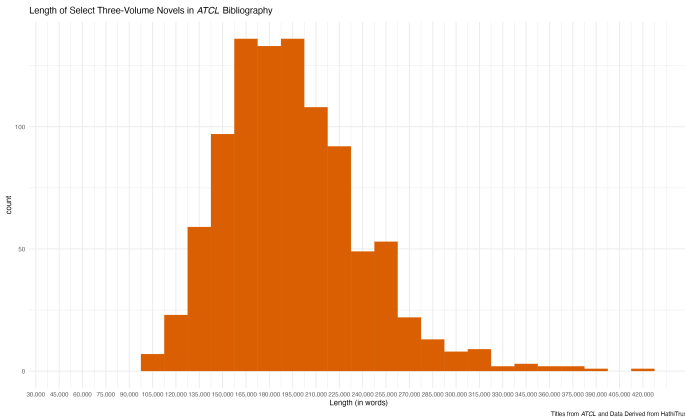Figure 1: Histogram of Lauterbach Book Length Data

Figure 2: Histogram of Book Length Data Derived from HathiTrust and ATCL
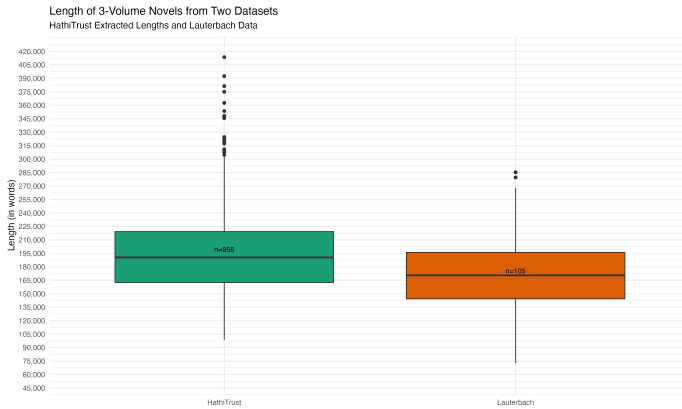
Figure 3: Box Plot Comparing Lauterbach Data with HathiTrust-Derived Data

would simply confirm that my method of tallying three-volume novels from HathiTrust data was accurate. So when I saw this box plot (and for quite a while afterwards) my heart sunk. There is a significant difference between the means of these two datasets (`Lauterbach mean=171038, HathiTrust mean=194818`)Welch's t-test returns a p-value of `0.000003797`; which suggests the between these two populations is not mere chance. which should ostensibly be describing the same general phenomenon.

When I first tried to explain this discrepancy, I noted that Lauterbach and Lauterbach's sense of the three-volume novel includes a number of pre-Victorian texts; it includes, for instance, *Frankenstein* (published in three-volumes in 1818) as well as fifteen other texts published before 1827 (the earliest date of the *ATCL*-derived, HathiTrust data). The Lauterbachs' data include novels that just happen to have been published in three-volumes, not necessarily "three-volume" novels of the sort produced by the rise of the lending libraries. So I removed them from the dataset, assuming that pre-Victorian three-volumes were shorter (*Frankenstein* is the short-
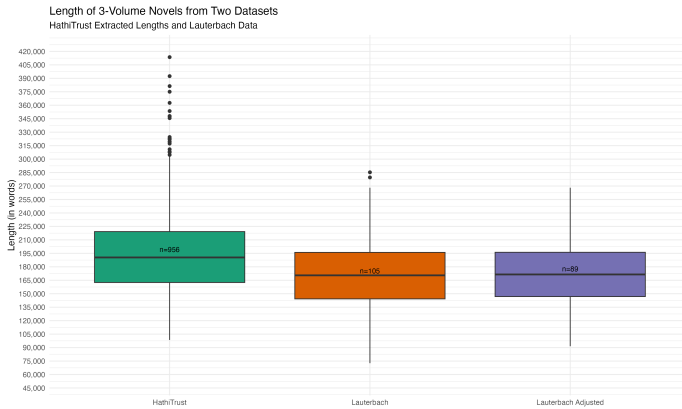
Figure 4: Box Plot Comparing Lauterbach Data with Original HathiTrust-Derived Data

est novel in the dataset!), and that these novels were reducing the Lauterbach's data.

But removing those texts made no real difference (the reduced set is "Lauterbach Adjusted" in the following boxplot, `mean length 173,284 words`).

Perhaps, dear reader, you noted the blithe way above I glossed over a rather important detail in how I counted the length of the HathiTrust texts in the code above. If you noticed it, congratulations! Sadly, I did not (without a lot of frustration). My counts were sums of *tokens*. That includes tokens representing *punctuation*.

I revised my code to not count tokens that had been tagged with the following part of speech tags: `'$', '-RRB-', '-LRB-', ',', '.', "''", ':', '``'`[3] With that data, I got the following boxplot.

Here the median of the HT data is now 159,814 words. A Welch's

---

[3]I actually don't know what some of these tags (`-RRB-` and `-LRB-`) represent. I haven't seem this documented anywhere, and don't see such tags in other descriptions of Penn's Treebank tags.
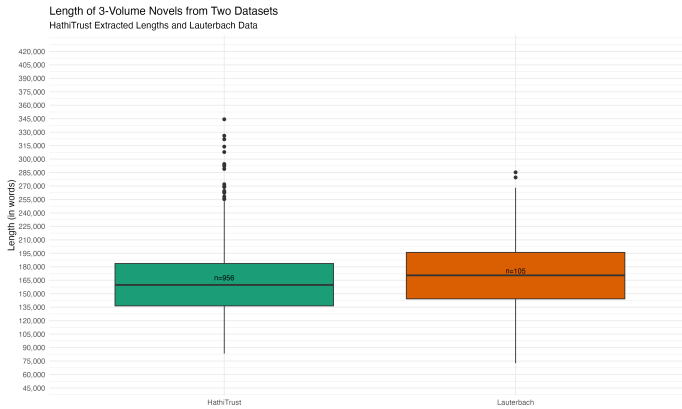
Figure 5: Box Plot Comparing Lauterbach Data with Adjusted HathiTrust-Derived Data

T-Test between the HT and Lauterbach data results in a p-value of 0.08, which suggests the difference may not be significant, for whatever that's worth.

While there is still a detectable difference, it is smaller, and seems consistent with what we see if we compare the number of *pages* reported by Lauterbach and in the *ATCL* titles detected in HathiTrust:

I suspect that it would make sense to further fine-tune which tokens are counted, but this adjustment satisfied me that my data were telling the same basic story as Lauterbach's (and so would provide a safe basis for additional, more sophisticated analyses).

Length Over Time

Based on this comparison with Lauterbach's data, we can be at least somewhat confident in other uses we put to the data we extract from HT using titles from *ATCL*. The more interesting uses will use the word count data to make inferences about the narratives themselves. Before I end this post, however, it is worth looking at the length of the three volume over time. So far, my data has only con-
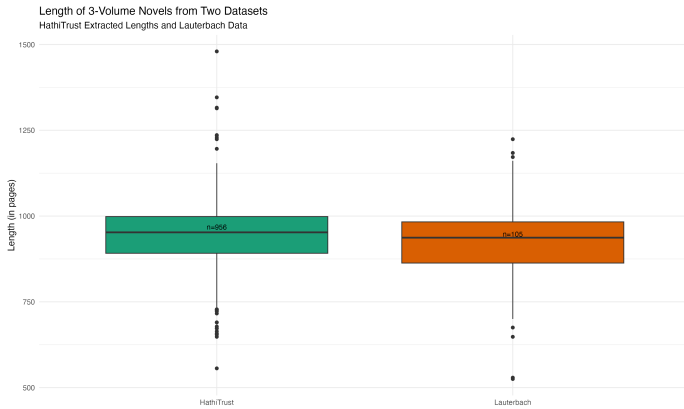
Figure 6: Box Plot Comparing Number of Pages in Three-Volume Novels, Lauterbach Data with HathiTrust-Derived Data

firmed the story told Lauterbach, Gaskell, and Bassett. Combining more titles (provided by Bassett's bibliography) with an analysis of length (akin to Lauterbach's), we can look at something none of those critics have considered—what happened to the length of the three-volume novel over its history.

Three-volume novels get shorter over the 70-year period represented by this data. The trend is slight, and easier to see in these histograms, split into three periods.

Over its history, the three-volume novel got shorter by almost 30,000 words, or about 90 pages. That seems at least a little interesting! More interesting things, though, can done with this data—in a future post.

Works Cited

Bassett, Troy J. *At the Circulating Library.* At the Circulating Library: A Database of Victorian Fiction, 1837—1901, https://www.victorianresearch.org/atcl/.

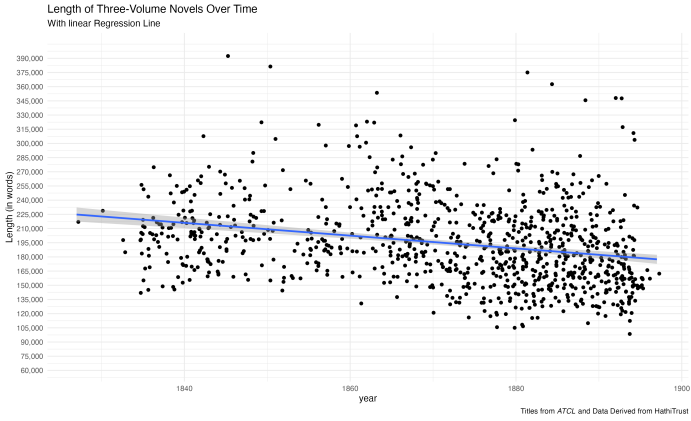Gaskell, Philip. *A New Introduction to Bibliography.* Reprinted with

Figure 7: Scatter Plot Showing Length of Three-Volume Novels (in words) Over time
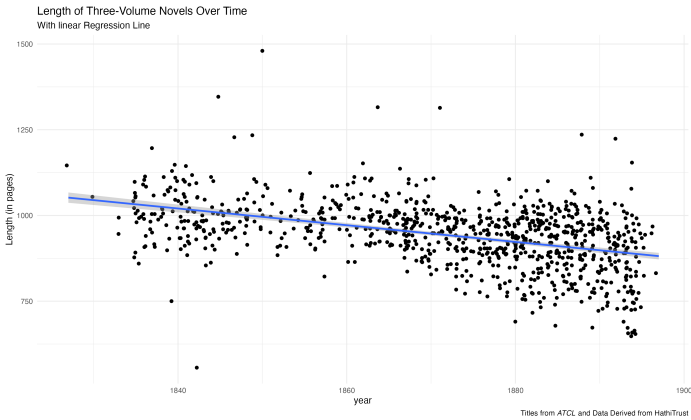

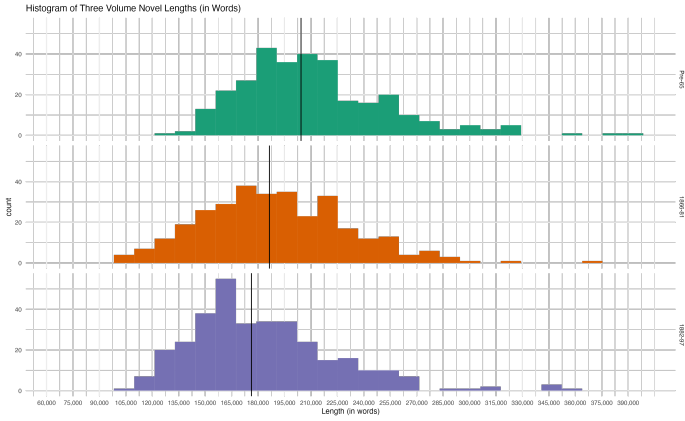Figure 8: Scatter Plot Showing Length of Three-Volume Novels (in words) Over time

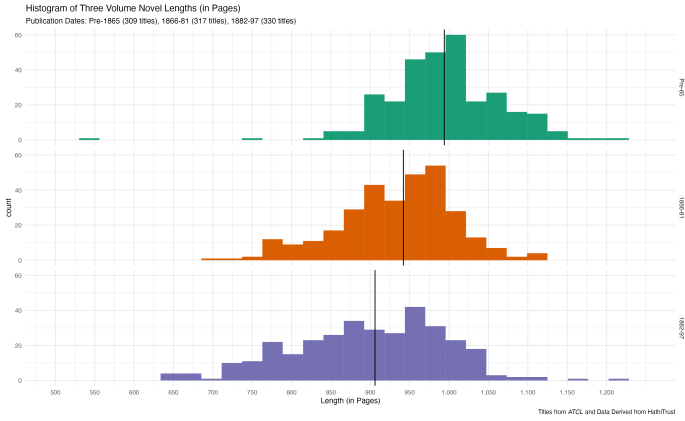Figure 9: Histogram of Novel Lengths (in words) Divided into Three periods

Figure 10: Histogram of Novel Lengths (in pages) Divided into Three periods

corrections in 1995, Oak Knoll Press, 2007.

"Guinea." *OED*, https://www.oed.com/dictionary/guinea_n?tab= meaning_and_use#2320395. Accessed 16 Aug. 2023.

Jackson, Gita. *Why Does Everything On Netflix Look Like That?* Vice, 15 Aug. 2022, https://www.vice.com/en/article/ake3j5/why- does-everything-on-netflix-look-like-that.

Lauterbach, Charles E., and Edward S. Lauterbach. "The Nineteenth Century Three-Volume Novel." *The Papers of the Bibliographical Society of America*, vol. 51, no. 4, 1957, pp. 263–302, https: //www.jstor.org/stable/24299448.

Wilson, Nicola. "Circulating Morals." *Prudes on the Prowl: Fiction and Obscenity in England, 1850 to the Present Day*, edited by David Bradshaw and Rachel Potter, 1st ed, Oxford University Press, 2013, pp. 52–70.