

Cormac McCarthy Oeuvre Analysis

Jon Michael Stroh

Introduction

Cormac McCarthy, who passed away in 2023, is widely considered among the greatest novelists of the 20th and early 21st century. Stephen King praised him as “the greatest American novelist of my time”.¹ His novel *All the Pretty Horses* won the National Book Award in 1992, and *The Road* won the Pulitzer Prize in 2007.² The film-adaptation of his novel *No Country for Old Men* won the Oscar for Best Picture in 2007.³ In total, he published 12 novels between 1965 and 2022.

Many such as I have been entranced by McCarthy’s distinct prose. Often his novels follow the gruesome and grisly, yet with a poetic and entrancing style. He writes minimally with almost no punctuation, following the likes of Hemingway, but often with lyrical grandeur, like Faulkner, with striking, evocative, specific descriptions.

Academics, fans and followers have noticed a schism among his collection.⁴ His first four novels, all set in Appalachia, are considered hopeless, dark and bleak works.⁴ While his later novels, especially the border trilogy of *All the Pretty Horses*, *The Crossing*, and *Cities of the Plain*, which received more public support and financial success, have their moments of warmth and kindness, uncommon occurrences in McCarthy novels.

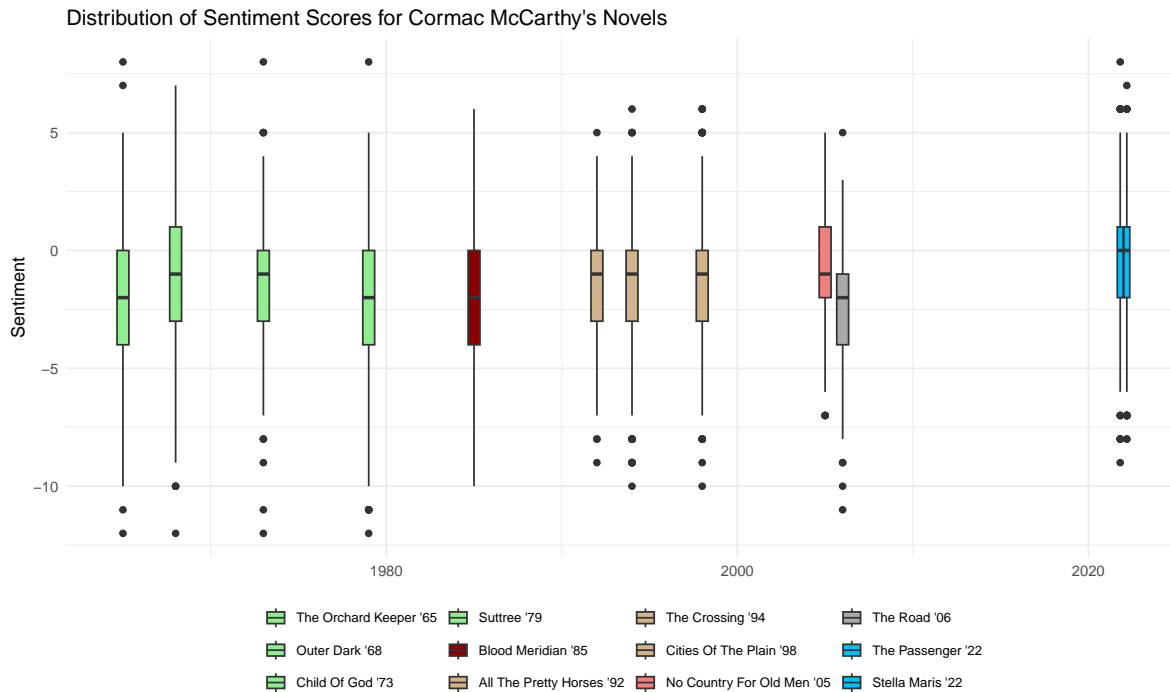
Using sentiment analysis on his novels, I sought to determine whether or not a divide is actually present. Did McCarthy’s writing, on average, get more positive over time as his commercial status improved?

The data includes McCarthy’s twelve novels, tediously scraped from the web in a myriad of formats – .txt, .epub and .pdf. The cleaning process, likewise tedious, depended on the file type. For each individual novel, I had to view the text and remove unwanted beginnings and endings. Additionally, for .epub files, which are eBooks, I had to remove new lines, tab lines, and occasionally page numbers from the text. To double-check, I compared the word counts of my files with known word count estimates for each novel. Finally, I sectioned the text of each novel into 700 character portions (without splitting words), as an analysis of Jane Austen’s novels did, to ready them for sentiment analysis.⁵

For sentiment analysis, I employed the syuzhet package. The package itself includes a number of lexicons and dictionaries for calculating sentiment in text – AFINN from Finn Årup Nielsen, Bing from Bing Liu and collaborators, and NRC from Saif Mohammad and Peter Turney.⁶ After investigating all three, I settled on Bing for my primary analysis. The AFINN score has a much higher variance than the others. And the NRC score tends to skew more positively compared to the others. Thus, Bing as the preferred medium. Also, in an analysis of Jane Austen’s novels for which I found inspiration, Silge notes Bing sentiment appears to find longer stretches of similar text, which resonates with the type of analysis I wish to conduct, analyzing the continuous sentiment of his novels.⁵ However, I will test the robustness of the analysis on the other two scoring systems. After running the functions on the data, I combined all 700 character sections of each novel and their sentiment scores into a single dataframe:

	bing	afinn	nrc	title	year
1	-1	-4	-3	The Orchard Keeper	1965
2	-8	-9	-1	The Orchard Keeper	1965
3	-5	-7	-4	The Orchard Keeper	1965
4	-4	0	-1	The Orchard Keeper	1965
5	-2	-3	-3	The Orchard Keeper	1965
6	3	3	1	The Orchard Keeper	1965

Next, I explored the data visually:



First, I notice the general negativity of Cormac’s novels, confirmed here in the initial look at sentiment. The mean sentiment of every novel falls below 0 (*The Passenger* and *Stella Maris*, the final two novels, fall just below). Second, interestingly enough, the sentiment for the *Border Trilogy* from ’92 through ’98 have very similar sentiments. And, the final sister novels, *The Passenger* and *Stella Maris*, also have mirroring sentiment distributions. It appears McCarthy deftly matches the sentiment distributions of related novels. Here, I observe some evidence for increasing positivity over time.

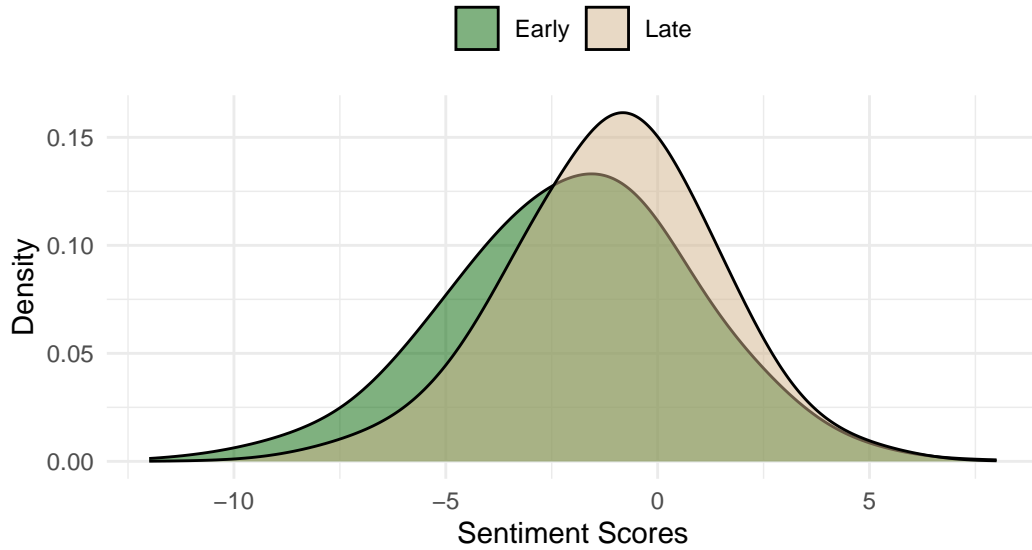
Methodology

To investigate the differences between McCarthy’s earlier and later works, I needed to define a line for such division. Utilizing academic opinion and my own knowledge around his collection (I’ve read all twelve novels), I decided to categorize his first five novels, beginning with *The Orchard Keeper* in 1965 and ending with *Blood Meridian* in 1985 as *early* period, and the final seven novels, beginning with *All the Pretty Horses* in 1992 and ending with sister novels *The Passenger* and *Stella Maris* in 2022 as *late*. The only significant point of contention was whether to classify *Blood Meridian* as early or late. Although set in the American southwest like many of his later novels, I decided on *Blood Meridian* as early for two reasons. First, the prose resembles his earlier style with its unorthodoxy and bold and beautiful disregard for convention.⁷ While his later novels have prose no less breathtaking, they heed, on some small level, to more conventional structure. Second, although *Blood Meridian* is now considered among the greatest novels of all time, it was McCarthy’s last novel before gaining popular appeal.⁸

I considered a number of options for testing the difference in the sentiment of McCarthy’s writing between his early and late period including a two sample t-test, Kolmogorov-Smirnov test, and bootstrapping techniques. The most straightforward approach, a t-test, initially concerned me, particularly on the fronts of total data and normality. First, I had concerns about the limit of just 12 novels as data. However, I realized – and I specify once again here – that I’m not focusing on testing if McCarthy’s novels as a whole got less bleak, but rather did his writing in general become less negative. This provides justification for dividing the novels into 700 character portions to conduct sentiment analysis on. This will value longer novels more, in line with my research goal, because longer novels provide a larger sample of McCarthy’s writing and thus should be valued more.

Second, I needed to confirm normality for the sentiment of McCarthy’s writing for both early and late periods. Does McCarthy write, or does the dictionary capture, many sensationalistic sections, either positive or negative? Or does, generally, his writing mirror the normal distribution?

Density of Bing Sentiment for Early vs. Late McCarthy



Again, Early is '85 and before

I've already noticed the median negativity among his writing, but overall, both his early and late career writing follow the normal distribution.

Table 1: Variance of Sentiment Scores by Period

Period	Variance
Early	8.431
Late	6.186

Something to monitor is the slightly greater variance in sentiment for early period writing. However, I concluded a Welch's (for the difference in variance) t-test was appropriate for testing whether there's a difference between the mean sentiments of his early and late periods.

Results

Table 2: Welch's T-Test Results for McCarthy's Early vs Late Period Bing Sentiment

	Value
Early Period Mean	-2.016
Late Period Mean	-1.091
Difference 95% CI	(-1.046, -0.804)

	Value
T Statistic	-14.981
Degrees of Freedom	6692.794
P-Value	<2e-16

The sentiment scores derived from the novels indicate a significant difference between the early and late periods of McCarthy’s writing with an extremely low p-value of <2e-16. Specifically, the mean sentiment score for the early period was -2.016, while for the late period, it was -1.091. This suggests a shift towards less negative or more neutral sentiment in the writing of his later works. The confidence interval for the difference in means between the two periods was (-1.046, -0.804), not encompassing zero, which further supports the conclusion that there is a significant difference in sentiment between the two periods.

The test results support the hypothesis that Cormac McCarthy’s writing underwent a significant shift in sentiment over time, with his later, more commercially successful, novels characterized by more positive sentiment, on average, compared to his earlier works. Again, both his early and later novels have negative sentiment overall.

Upon retesting with the AFINN and NRC sentiments, I return significant results with the same conclusion, also with an extremely low p-value (see the Appendix). Therefore, regardless of the differences in the individual metrics, the three sentiment lexicons all appear to be capturing the same trend which provides evidence for the robustness of this approach.

Table 3: Mean Bing Sentiment

Title	Year	Mean
The Orchard Keeper	1965	-1.80
Outer Dark	1968	-1.19
Child Of God	1973	-1.71
Suttree	1979	-2.25
Blood Meridian	1985	-2.26
All The Pretty Horses	1992	-1.13
The Crossing	1994	-1.64
Cities Of The Plain	1998	-1.05
No Country For Old Men	2005	-0.61
The Road	2006	-2.19
Stella Maris	2022	-0.30
The Passenger	2022	-0.53

Of course, there are exceptions to this trend. *The Road*, published in 2006, has the third lowest mean sentiment among all twelve novels. And *it is* an incredibly heartrending, dismal novel,

yet had significant commercial success, won the Pulitzer Prize, and was selected by Oprah's book club.⁹ With this in mind, I do imply with this analysis that the more positive sentiment present in McCarthy's later writings lead to or caused commercial success.

Discussion

The analysis suggests that McCarthy's later works, starting from *All The Pretty Horses* through *Stella Maris*, may manifest a relative softening in the harshness of tone and description that characterized his early works. This shift could be associated with a variety of factors including changes in personal philosophy, maturation of style, or even market dynamics influencing literary content. Notably, this transition does not necessarily imply that his later works are upbeat or devoid of the characteristic McCarthy bleakness; rather, they might integrate more moments of warmth or redemption that were less apparent in his earlier narratives.

There are several potential limitations to this analysis. First, simply, is the limitation of sentiment analysis and the three lexicons I employed. Sentiment analysis tools, while powerful, cannot fully capture the nuanced emotional and thematic layers that a human reader or literary critic might discern. Second, I often reference the overall negative sentiment of McCarthy's work captured by the analysis. However, this baseline assessment, in fact, comes not from the data but from literary analysis of his work that I then associated with the data.¹⁰ So, without comparing these findings against a broader dataset of contemporary literature beyond McCarthy, it's challenging to conclude definitively whether McCarthy's work is uniquely negative in sentiment or simply reflects broader literary trends. Perhaps, novels generally skew negative as captivating conflict is often, in fact, negative.

And so, further research could benefit from a comparative analysis involving a broader spectrum of authors and genres to contextualize McCarthy's sentiment trajectory against wider literary trends. This would help ascertain if the observed shifts in sentiment are idiosyncratic to McCarthy or part of a larger shift in literary narrative styles.

Additionally, an intriguing avenue for further study could involve correlating the sentiment of McCarthy's novels with their critical and commercial success. My preliminary observations suggest novels like *Suttree*, *Blood Meridian*, and *The Road*, which are among his most critically acclaimed yet also exhibit the lowest sentiment scores, may indicate that darker tones could correlate with higher literary acclaim in McCarthy's case.¹¹ Expanding this analysis to include reader reviews, literary awards and sales data could provide deeper insights into the relationship between narrative sentiment and literary success.

This analysis underscores the evolution in Cormac McCarthy's literary sentiment over decades, suggesting a nuanced yet statistically significant shift towards less negative expressions in his later works. While the findings are compelling within the scope of the conducted analysis, they invite broader literary and cultural studies to fully unravel the complexities of McCarthy's literary legacy.

Citations

1. Steger, J. (2023, June 14). The cormac mccarthy novels you should read. The Sydney Morning Herald. <https://www.smh.com.au/culture/books/the-five-cormac-mccarthy-novels-you-should-read-20230614-p5dggp.html>
2. Cormac McCarthy. Penguin Random House Canada. <https://www.penguinrandomhouse.ca/authors/>
3. Tapp, T. (2023, June 13). Cormac McCarthy Dies: Pulitzer prize-winning author of “no country for old men,” “The road” & “blood meridian” was 89. Deadline. <https://deadline.com/2023/06/cormac-mccarthy-dead-author-no-county-for-old-men-the-road-1235416172/>
4. Garner, D. (2023, June 13). Cormac McCarthy, novelist of a darker america, is dead at 89. The New York Times. <https://www.nytimes.com/2023/06/13/books/cormac-mccarthy-dead.html> /
5. If I loved natural language processing less, I might be able to talk about it more. Julia Silge. (2016, March 18). <https://juliasilge.com/blog/if-i-loved-nlp-less/>
6. Mjockers. (n.d.). Mjockers/Syuzhet: An R package for the extraction of sentiment and sentiment-based plot arcs from text. GitHub. <https://github.com/mjockers/syuzhet>
7. J. P. (n.d.). What the style of Blood Meridian Makes impossible. Daniel David Wallace. <https://danieldavidwallace.com/2014/06/24/what-the-style-of-blood-meridian-makes-impossible/>
8. Sinykin, D. (2023, June 19). Cormac McCarthy had a remarkable literary career. it could never happen now. The New York Times. <https://www.nytimes.com/2023/06/19/opinion/cormac-mccarthy-publishing.html#:~:text=>
9. Person. (2007, March 28). Your reader’s guide to the road. Oprah.com. <https://www.oprah.com/oprahsbookclub/readers-guide-to-the-road-by-cormac-mccarthy/all>
10. Magazine, S. (2023, June 15). Behind the bleak, beautiful stories of Cormac McCarthy. Smithsonian.com. <https://www.smithsonianmag.com/smart-news/the-stories-that-cormac-mccarthy-left-behind-180982380/>
11. Goodreads. (n.d.). Best of Cormac McCarthy (18 books). Goodreads.

Appendix

Density of AFINN Sentiment for Early vs. Late McCarthy

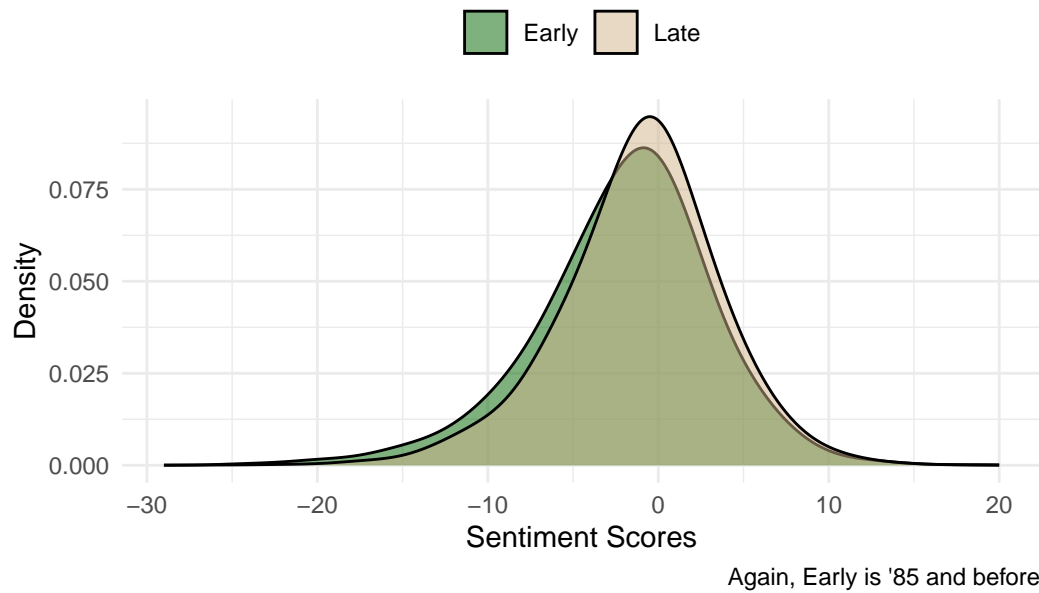


Table 4: Welch's T-Test Results for McCarthy's Early vs Late Period AFINN Sentiment

	Value
Early Period Mean	-2.027
Late Period Mean	-1.095
Difference 95% CI	(-1.153, -0.712)
T Statistic	-8.282
Degrees of Freedom	6986.803
P-Value	<2e-16

Density of NRC Sentiment for Early vs. Late McCarthy



Table 5: Welch's T-Test Results for McCarthy's Early vs Late Period NRC Sentiment

	Value
Early Period Mean	-1.481
Late Period Mean	-0.406
Difference 95% CI	(-1.197, -0.951)
T Statistic	-17.138
Degrees of Freedom	7244.048
P-Value	<2e-16