Faulkner and the Politics of the Long Sentence

Johannes Burgers

## Introduction

In the earliest reviews, Faulkner’s style was divisive: high art for some, bad writing for others. This friction has created a rich body of scholarship on the form of Faulkner’s writing generally and his use of punctuation in particular. What has eluded critics, however, is a view of his writing as a whole. New computational techniques make it possible to sketch a more detailed portrait of his use of punctuation throughout his career.[[1]](#footnote-20) This data can be coupled with the character, location, and event data available in the *Digital Yoknapatawpha* (*DY*) database. Combined, this data can be used to understand not only how often Faulkner makes particular choices about punctuation, but also the relationship between punctuation choices and the characters present or mentioned in those particular sentences. Doing this confirms suspicions about specific punctuation patterns in Faulkner by earlier scholars, and furthers the conversation about who gets represented in Faulkner and how.

In the main, punctuation in Faulkner is a means to achieve his aesthetic of the long sentence. This aesthetic tries to capture a person’s past, present, and possible future in one moment in time, and thereby collapses the neat temporal distinctions that cast the impression of a stable, singular self moving through history. Instead, the continuous persistence of the past and possible futures of the long sentence suggest a self that is always in the process of being constructed. Though ostensibly representative of the human condition writ large, the aggregated data tell a different story from Faulkner’s own comments on his writing. Sentence length and the concomitant experimentation with punctuation that makes the long sentence possible are strongly correlated to the presence of upper class White male characters. Though each individual long sentence may be thematically unique and contain characters from a multitude of backgrounds, the representation of these themes and characters is invariably modulated by the presence of an upper class White men. Quite paradoxically then, Faulkner’s radical experimentation with punctuation often serves to centre the very upper class White male perspective it seeks to trouble.

## The Sum of the Past

Faulkner’s style has long been a central concern of critics and scholars. It has not always been universally well-received, one critic, writing of *Absalom, Absalom!,* remarks that, “When a narrative sentence has to have as many as three parentheses identifying the reference of pronouns, it signifies mere bad writing and can be justified by no psychological or esthetic principle whatever” (DeVoto 147). Other contemporaneous critics did find merit in Faulkner’s long sentences and saw them as a “microcosm” (Aiken 47), “self-contained world” (Van O’Connor 83), and tool for “compressing the greatest possible amount of time, or life, or motion, into the smallest possible space, in order to condense and stop it for contemplation” (Adams 107). As one critic puts it, Faulkner’s sentences are the “nearest that language as the instrument of fiction can come to the instantantaneous complexities of consciousness itself” (Beck 47). Faulkner himself offered a similar view. When asked about his use of long sentences at the Virginia Colleges Conference in 1957, he claims that there is no such thing as “was” because the past is part of “every man, every woman, at every moment”(Faulkner, *Virginia Colleges Conference, Tape 1*). This aesthetic attempts to get “his [a character in a story] past and possibly his future into the instant in which he does something”(Faulkner, *Virginia Colleges Conference, Tape 2*). On the surface, this has a universal appeal: men, women, and everyone is the “sum” of their past (Faulkner, *Virginia Colleges Conference, Tape 1*). It suggests that the long sentence is the collision point between past, present, and future and the ineluctable entanglement of all three as a fundamental human condition.

Later critics have subsumed the aesthetic of the long sentence into a general discussion of Faulkner’s style. In ‘Faulkner’s Formalisms’ Sebastian Fanzun provides a nuanced and comprehensive overview of this genealogy. Fanzun recognizes a basic polarity in approaches between the New Criticism exemplified by Cleanth Brooks that tries to unify the texts into coherence, and a post-structuralist turn initiated by John T. Matthews’s highly influential *The Play of Faulkner’s Language*, which argues for the impossibility of such an “organic unity” (Fanzun 23). The disunity in Faulkner’s style Matthews introduces in part helped create the opportunity for different perspectives within Faulkner to emerge. There remains, though, a lingering duality in the relationship between the radical style and the conservative politics of Faulkner’s works. One the one hand, language in Faulkner has “the power to say everything” (Lockyer 68), is a “regenerative force” (Olsen 110), and, in the case of Rosa Coldfield, is filled with “queer potentiality” (Thomas 49). On the other hand, the language is fundamentally limited and limiting, where Faulkner’s ways of writing about a Black character, Lucas Beauchamp, are “his ways of speaking his own racial identity” (Weinstein 67), that he exercises a “politics of quietness” in *Light in August* that unifies his “racism and anticommunism” (Meyerson and Neilson 39), and that his representations of Native Americans echo his own and his culture’s “anxieties about white masculinity” (Parker 96). Computational analysis cuts these tensions by demonstrating the simultaneous presence of continuous formal experimentation that resists a unified view of Faulkner’s writing and, indeed, coherence at the sentence level, while also highlighting the unifying vision that underpins all of this experimentation.

## Computing Faulkner’s Corpus

Computational analysis of Faulkner’s writing presents some thorny issues that are actually quite productive when considering his use of punctuation. The most significant issue is that there is not single, settled upon edition of his texts. Noel Polk famously edited “corrected” versions, but even he admits they are far from the “definitive” (Polk 6). In part, this is because Faulkner was himself inconsistent with his approach punctuation. A salient example is the Benjy section in *The Sound and the Fury*, which Faulkner retyped three times using a different punctuation system(Polk 14). These permutations are evidence of Faulkner’s constant experimentation with punctuation and his style more generally throughout his career. Still, Polk’s own modesty about his accomplishment notwithstanding, this paper uses his “corrected” editions for the corpus where possible as these also align with the texts used for creating the *Digital Yoknapatawpha* database.

Beyond the lack of a “definitive” edition another challenge for computational analysis is the internal coherence of the textual corpus. Faulkner’s writing is very heterogeneous, consisting of novels, short stories, poetry, film scripts, and a text that hovers between a stage play and a novel: *Requiem for a Nun*. As these texts all represent different types of writing, only novels and short stories were considered. Of these, *Requiem for a Nun* had to be excluded as well because it was radically different from all of Faulkner’s other prose.

Furthermore, Faulkner’s corpus also represents a unique computational challenge due to its intertextuality. There are the 14 novels and 54 short-stories that take place in his apocryphal Yoknapatawpha County. Of these short stories, many of which were originally published in popular magazines, about 5% was were incorporated wholesale or in part into novels such as *The Unvanquished* (1938), *The Hamlet* (1940), and *Go Down, Moses* (1942), resulting in .4% of the text being duplicated verbatim across the corpus. This intertextuality does not follow a neat timeline. Some works written early in his career are only published much later, like “Afternoon of a Cow” which was published forty years after he first sketched out the idea (Faulkner, *Uncollected Stories of William Faulkner* 703). Meanwhile, two substantially different versions of “The Bear”, one in The *Saturday Evening Post* the other in *Go Down, Moses,* are published within days of each other (Corrigan and Padgett). This makes it difficult to establish a concrete timeline of his stylistic development, because composition date and publication date can vary widely. It also means that linguistic patterns that appear in one text may in a certain sense *belong* to a previous text.

In sum, this paper is an analysis of some but not all of Faulkner’s writing, and it is difficult to establish a concrete picture of linguistic patterns due to a variance in editions, textual heterogeneity, and overlap between texts that may be years apart in publication. These complications should militate against making sweeping generalizations about Faulkner’s writing. What is true in some instances may not be true in all instances.

## Faulkner’s Boundless Sentences

Perhaps the most consistent sweeping generalization about Faulkner is that he writes long sentences. To be sure, it hardly requires computational analysis to understand that some of Faulkner’s most famous works contain long sentences. For anyone who has ever taught Faulkner, it is perhaps the most common stumbling block and, indeed, complaint among students. Faulkner himself was asked about it no fewer than four times, when he lectured at at the University of Virginia (Faulkner, *Virginia Colleges Conference, Tape 1*; Faulkner, *Local and UVA Communities, Tape 1*; Faulkner, *Undergraduate Literature Class*; Faulkner, *English Department Faculty and Wives*). However, quantifying what readers might mean when they say Faulkner’s sentences are long is more complex.

The first issue is defining what constitutes a sentence. In terms of computation, most algorithms generally use some notion of a sentence boundary like a full-stop, exclamation mark, or question mark to establish a sentence (Gagolewski 47–48). This technique works well in most cases, but runs into quite a number of exceptions in Faulkner. For starters, Faulkner has a tendency to nest narratives within quoted material. For example, in *The Hamlet* (1940) much of the story is told by Ratliff, an itinerant sewing machine salesman. When he takes over as narrator, the perspective moves from third person omniscient to first person, and within this embedded narrative Ratliff quotes various characters. These quotations are themselves sentences, but also part of a longer, enframing sentence. Thus, while describing a scene between two antagonists, Ratliff says, “What will he say? What can he say except ‘All right. What do you aim to do?’” (Faulkner, *The Hamlet* 13). Using strict sentence boundaries, this is four sentences. This seems unsatisfactory as this is all part of the same completed thought. The two sentences “All right.” and “What do you aim to do?” are embedded within the larger quote started with “What can he say except.” Not accounting for sentence embedding in *The Hamlet* reduces the average sentence length. The exact opposite problem is true in *The Sound and the Fury*. Here Faulkner embeds sentences and phrases through italics but does not use any traditional sentence boundaries. This occurs frequently in the the Quentin section where the character’s internal narration is interrupted by memories. At one point, Quentin recalls Gerald Bland’s mother talking about her son’s looks, which is interjected by a memory of Quentin fantasizing about shooting his sister’s fiance, Herbert Head: “Telling us about Gerald’s women in a *Quentin has shot Herbert he shot his voice through the floor of Caddy’s room* tone of smug approbation” (Faulkner, *The Sound and the Fury* 105). The main sentence is: “Telling us about Gerald’s women in a tone of smug approbation.” This is interrupted by the clause “Quentin has shot Herbert,” and possibly another sentence or a part of one, “he shot his voice through the floor of Caddy’s room.” There is no unambiguous way to determine if the two phrases should be seen as part of the same italicized sentence or if they should be read separately. There are valid interpretations for both. Similarly, while it would be computationally possible to excise the italicized text, this distorts the fact that these italics quite deliberately interrupt and extend the sentence because there is no end-sentence punctuation. It would also mean going against Faulkner’s explicit wishes to have the text presented in this way (Polk 13-14). Consequently, the sentence in *The Hamlet* (17 words) and the sentence in *The Sound and the Fury* (25 words) are roughly the same length, but in the tabulation the former is significantly shorter than the latter. This inconsistency is a feature of the entire corpus, because Faulkner varied the way in which creates long sentences throughout his career.[[2]](#footnote-24) Sentence length is therefore a rather crude and imperfect measure to understand punctuation in Faulkner, but it makes an interesting starting point for investigation.

Going by the imperfect measure of sentence length, Faulkner’s writing is not exceptional compared to other writing at the time. At 16.77 words-per-sentence (wps), Faulkner’s sentences are only slightly longer than the average 16 wps for a text in the Corpus of English Novels, which covers writing from 1881 - 1922 (Ihrmark and Nilsson 79) and longer still than the 13 wps around the 1930s (Rudnicka). Nor is the slightly higher than average sentence length a feature of all his works. The *Sound and the Fury* is notoriously complex. Nevertheless, at only 10 wps, it uses fewer words on average than Hemingway’s *The Old Man and the Sea*, which uses 13 wps(Ihrmark and Nilsson 82). Despite this counter-intuitive disparity, Faulkner’s more verbose style could never be confused with Hemingway’s trademark concision. In part, this has to do with the distribution of sentence lengths within each text. The *The Sound and the Fury* may have a lower words-per-sentence average than *The Old Man and the Sea*, but it also contains one sentence that counts over 4,000 words. Clearly, this is less succinct than Hemingway, and probably most other authors as well. These long sentences, it should be noted, are more of a statistical exception than the rule. The normal distribution for *The Sound and the Fury*, and most of Faulkner’s other texts is heavily right-skewed. Meaning that most sentences are of average length, but a few are much, much longer. In the case of the *Sound and the Fury* only about 13% of the sentences are longer than the mean corpus sentence length. The same can be said for Faulkner’s other texts, most of these (62%) have an average sentence length lower than the total average across the corpus. More succinctly, a small share of sentences in each work gives the impression of long sentences, and, similarly, within the corpus of Faulkner’s writing there are a minority of works that are above average in sentence length. Statistically, lengthy sentences are a local feature not a global phenomenon.

Nevertheless, lengthy sentences are seen as the hallmark of Faulkner’s writing. In part this may be because some of Faulkner’s most canonical works feature long sentences, *Go Down, Moses* (20), *Absalom, Absalom* (43), ‘A Rose for Emily’ (19), and ‘Barn Burning’ (19) These works are, in a sense, more Faulknerian than a relatively unknown early text like ‘Elly’ (9). Nevertheless, equally canonical works like ‘That Evening Sun’ (9), *Light in August* (15) , and *As I Lay Dying* (12) do not, on average, stand out as being particularly verbose relative to the the general corpus of texts written in English during this period. Despite these counter indications, it goes against a firmly established scholarly tradition based on Faulkner’s own statements and, indeed, common sense, to argue that lengthy sentences are not Faulkner’s particular metier. Instead, to understand why Faulkner’s sentences, for lack of a better word, *feel* long it is informative to look at their internal structure.

## Extending the Sentence: Faulkner’s Experiments in Medial Punctuation

The preceding discussion about sentence length raises an obvious question: What is a long sentence? Is it any sentence above the average sentence length of the corpus? If so, how much longer does it need to be to be consider long and not simply longer than usual? If a sentence is indeed long, is there a maximum length? At what point, does a string of words grammatically cease to be sentence? These questions can, in part, be answered by drawing a contrast with another modernist who inspired Faulkner and who was also famous for his long sentences: James Joyce. Joyce wrote some of the longest sentences in the English language in the Molly episode of *Ulysses,* but they are not long in the same way a sentence in Faulkner is long (Fargnoli and Golay 102). The Molly episode is essentially a string of syntactically coherent sentences with the punctuation removed (Attridge 545). Faulkner not only creates long strings of words, but they are rendered syntactically unconventional and challenging through his use of punctuation. Thus, whereas Joyce saw the intervention of punctuation as an infelicity to the ineluctable flow of language in Molly’s soliloquy,[[3]](#footnote-26) the sheer profusion of punctuation in Faulkner speaks to the endless way in which meaning branches, eddies, rejoins and ends in an arbitrary terminus only to recirculate again. Molly’s soliloquy is a stream, Faulkner’s writing is a watershed.

To understand how Faulkner lengthens his sentences it is useful to look at medial punctuation like: colons, parenthesis, ellipses. As earlier critics observed, Faulkner uses this punctuation in an unconventional way, and, in doing so, calls attention to the punctuation itself (Scott 93). Arguably, any number of texts serve as good examples, but there are three salient texts that deviate from Faulkner’s own writing statistically: *Absalom, Absalom!*, *Intruder in the Dust*, and ‘Miss Zilphia Gant.’

A statistical model of the punctuation in his short stories (n= 93) and novels (n= 18) can help determine which texts have punctuation at variance with the corpus more generally.[[4]](#footnote-27) The texts were broken down into sentences, and each mid-sentence punctuation mark was counted. Intuitively, sentence length and mid-sentence punctuation are positively correlated. As sentences get longer, there is a higher likelihood of finding mid-sentence punctuation. There are some functional limits to this correlation, however. Modern punctuation tends to function as a syntactical system that regulates the number of certain punctuation marks that can be used (Schou 198). Hence, when one sentence in *A Fable* (1954) has 53 commas, 9 semi-colons, 6 colons, 2 parentheses, and is over 750 words long (Faulkner, *A Fable* 253–55), it disrupts conventional understanding of how those punctuation marks should be used.

By taking the averages of the number of medial punctuation marks per sentence across all texts, it is possible to detect texts that deviate from the overall pattern. In general, phonological and syntactical phenomena tend to follow a power-law distribution (Newman 372; Sun and Wang 32); where the probability decreases exponentially with each frequency interval. Syntactically, this makes sense. In any given text, most sentences do not contain commas. If a sentence does contain a comma, it is far less likely that it would also contain a second comma, it is exponentially less likely to contain a third comma, and so forth.[[5]](#footnote-28) The curve for these distributions looks a lot like a very steep cliff protruding from a plain; the line drops from the y axis into a sharp turn to the right where it meets the x axis and extends in a long tail. If Faulkner did not experiment a similar distribution would be present, and there would be little variance in the number of punctuation marks. This is not borne out by the data. Plotting the distributions across all punctuation marks reveals an exponential distribution best described as chi-squared with three degrees of freedom (). Visually, this curve looks a bit like a slide on a playground: a steep ladder on the left-hand side with a long-curve that never quite reaches the ground on the right. This curve indicates that many texts are within the same range, but a few texts have a very high relative frequency of specific punctuation marks.

Determining which values are outside the expected range and therefore outliers is a matter of some interpretation. A variety of tests revealed that for each mid-sentence punctuation mark different texts were outliers, which suggests an inconsistent use of mid-sentence punctuation.[[6]](#footnote-29) This variance does not appear to be strongly time dependent. There were no distinct phases to his punctuation experiments, and he does not, for example, use parenthesis heavily in the beginning of his career and semi-colons more extensively later in his career. Quite the opposite is true. Individual works stand out as having a substantially higher frequency of certain mid-sentence punctuation marks in comparison to others, and punctuation patterns that are the hallmark of some works are entirely absent in others. The reason why this difference is important is because these mid-sentence punctuation marks are a vehicle for creating long sentences, and function as a statistical signature for particular works. Three texts provide useful insights. *Absalom, Absalom*, and *Intruder in the Dust*, and ‘Miss Zilphia Gant’ were all marked as statistical outliers.

As one of Faulkner’s most challenging texts, it is unsurprising that the punctuation patterns in *Absalom, Absalom!* diverge from the corpus. Averaging around 43 wps, the text is notable for its prolific use of parentheses. With this technique, sentences are interrupted by parenthetical statements leading to sentences of extraordinary length. In one sentence stretching over 1,000 words, Faulkner uses nine parenthetical statements as a running commentary on the main sentence (Faulkner, *Absalom, Absalom!* 148–50). Some of these parentheticals are themselves several lines. All of this is complicated by the fact that this 1,000 word sentence appears in Chapter 6. The majority of this chapter is one long parenthetical statement. As a result, the parentheses in this sentence are actually nested within surrounding parentheses. More plainly, the sentence is an aside to an aside. Beyond this particular sentence, the nesting continues throughout Chapter 6 and at some point the the word Quentin is nested 4 levels deep (Faulkner, *Absalom, Absalom!* 168). That is, as an aside to an aside to an aside. There are so many parentheses that it appears that Faulkner or possibly his editor lost track of them, because the chapter has an uneven number of parentheses and is missing a final closing parenthesis. Not only, then, are Faulkner’s sentences very long in *Absalom, Absalom!* they are also, in a sense, deep. Each parenthetical statement adds more commentary and context, which itself requires more commentary and context. Therefore it is fitting that peeling away all of these parentheses like layers of an onion reveals the putative narrator: Quentin. The tale is ultimately about the teller.

In a slightly different fashion, Faulkner instrumentalizes colons in *Intruder in the Dust* to create sentence depth. To give an extreme example, one sentence in *Intruder in the Dust* is over 1,600 words long and features eighteen colons. The sentence deploys colons as a clausal adjunct. Geoffrey Nunberg labels this phenomenon colon-expansion, whereby the content following the colon expands or elaborates on the preceding clause (Nunberg, *The Linguistics of Punctuation* 30). Importantly, he points out that there are only two constraints to a colon-expansion. First, there is a semantic limit to the extent to which something can be elaborated, and, two, colon-expansions cannot themselves contain other colon-expansions (Nunberg, *The Linguistics of Punctuation* 30–31). Faulkner violates the former by dint of the latter. Each sentence dilates in meaning as he adds on more clausal adjuncts and elaborations. While it is impractical to quote Faulkner’s multi-page sentence in its entirety, the opening sequence captures this effect: “Because he was free : in bed : in the cool familiar room…” (Faulkner, *Intruder in the Dust* 34). With each colon more details about the narrative situation are revealed. The choice for concatenating clauses with a colon and not simply writing new sentences, appears to be similar to the depth effect created through parentheticals in *Absalom, Absalom!*. With each new colon the reader is tasked with processing a new piece of information that falls within the scope of one sentence. The narrative does not move forward but instead moves deeper.

It does not require inferential statistics to understand that *Absalom, Absalom!* and *Intruder in the Dust* have unusual punctuation patterns. One surprising statistical outlier is ‘Miss Zilphia Gant’, which has the highest relative frequency of ellipses in the corpus. It sits in the middle of a cluster of stories from 1927-1934 that average a higher number of ellipses per sentence than other Faulkner texts. To mark this as a general feature of Faulkner’s style during this period would be a mistake, however; as canonical works like *The Sound and the Fury*, *As I Lay Dying*, and *Light in August*, all written around this time, do not feature nearly as many ellipses. This was simply one of many punctuation marks he was experimenting with in some in some specific contexts. ‘Miss Zilphia Gant’ is perhaps the most extreme example. In this relatively obscure story, Miss Zilphia grows up a virtual captive of her overbearing mother. What makes the story unique is that its use of three dots is two-fold. Sometimes this indicates a traditional ellipse, at other times they function more like dashes that enclose an apposition (Railton). To complicate matters, in the early drafts, Faulkner, conversely, uses dashes as both traditional dashes, but also as ellipses. The effect is that the punctuation marks becomes a site of ellision as well an attentuation that creates a myriad of possibilities for the unsaid. For example, after Miss Zilphia finally marries a man, she still feels the urge to return to her mother. In the early typescript, her husband pleads with her, “if you go now - - - - - - Zilphy” (Faulkner, *"Miss Zilphia Gant," Bound Typescript 18 p.*). In the published version this becomes “If you go in there . . . Zilphy” (Faulkner, *Uncollected Stories of William Faulkner* 377). Stephen Railton speculates that this change was likely made by the editor of the Book Club of Texas, which first published this story (Railton). Whatever the reason for the final version, the multiple drafts all attest to the fact that Faulkner was experimenting with using entirely new forms of punctuation that could defer and displace the full meaning of the sentence to an unknowable place behind the ellipse.

The different techniques Faulkner uses to achieve his long sentences show a sustained commitment to an artistic vision he expressed throughout his career. Far from limiting experimentation to a particular phase or type of writing, his punctuation speaks to a constant desire to capture the long sentence. While Faulkner was certainly aspiring to capture a universal human condition, when these long sentences appear is not as universal as it might first seem. As a reader, it would be impossible to keep track of which characters appear in which long sentences. It is possible to capture this computationally, doing so reveals a stark fact: the long sentence appears to be reserved for only some people in Yoknapatawpha county.

## The Demographics of Punctuation

Understanding how the long sentence may or may not apply to certain characters requires knowing what type of characters are in what specific texts. Fortunately, the *Digital Yoknapatawpha* (DY) project has at least some of this data available. Started in 2012 with the mission to encode every character, location, and event in Faulkner’s Yoknapatawpha fictions, *DY* represents a rich database of every character in every event in every text that takes place in Yoknapatawpha. The database has been meticulously built by over thirty Faulkner experts through a strict peer-review process that has seen each of the thousands of records receive multiple editorial passes. The database focuses exclusively on the 14 novels and 54 short stories that take place in Yoknapatawpha, and excludes roughly 30% of the Faulkner’s writing in novels and short stories. Nonetheless, the database still represents the majority of Faulkner’s writing. In *DY* each event is any time a character or group of characters is present or mentioned at a specific location for a limited duration of time. Each character, in turn, is identified by different attributes including: name, race, class, and gender. By calculating what types of characters occur in each event, it is possible to build a demographic frequency model on how often and where characters occur in the corpus.

Based on the demographic frequency model, the world of Yoknapatawpha is heavily skewed towards Whites (79%) and men (80%). The distributions are also heavily left-skewed, meaning that the overwhelming majority of stories feature Whites and men, and in only very few stories do they not represent the majority of characters. Given that Faulkner was himself a White male, this is not entirely surprising. These measures give a rough indication as to the composition of demographics within the corpus. Race, class, and gender can also be linked to one another to get a better sense of the finer distinctions between characters.[[7]](#footnote-31) For example, upper class White males constitute around 24% of the characters who appear, whereas for enslaved Black women this number is a paltry 1%. Clearly, there is a stark disparity in terms of who gets represented.

These two data models, the punctuation and characters, can be combined to gain insight into the relationships between characters and punctuation. As with all such inferential models, the result are very much dependent on the underlying assumptions about the data. Suffice to say that each statistical analysis sheds light on possible relationships, but there is not one single test that provides a tidy answer. Given this, relationships between characters and punctuation can be imagined at three different scales: book, character type, and individual characters. In the first, the total character composition of the book is taken and compared to sentence length, in the second the character types in each event are compared to sentence length, and in the third each individual character is compared to sentence length. As the tests get more granular, there are more data points, and generally, there is a higher confidence that the pattern is not random.

### Texts

These demographic distributions per work can be matched to the punctuation distributions. At the level of the individual text, the relative frequency of character types that occur in the text can be compared to sentence length. A powerful way of doing this is through a Pearson correlation, which tests whether two numbers rise or fall together (), and whether this pattern is not random ().[[8]](#footnote-32) The Pearson correlation between sentence length and the race, class, and gender composite suggests a medium to strong positive correlation between sentence length and four types of characters: Enslaved women who have both Black and White ancestry ( = 0.535), upper class men who have both Black and White ancestry ( = 0.513), Men who have both Black and White ancestry and whose class is indeterminable ( =0.53), and upper class women who have both White and Black ancestry ( = 0.334). Needless to say, in Faulkner’s highly stratified social world, these characters are anomalous. If short stories are removed from the data set, this trend becomes even more pronounced. Indicating that this is more of a feature of the novels than the corpus as a whole. By and large, this correlation highlights that novels that deal with the transgression of the South’s strict race and boundaries have a higher tendency to use long sentences: *Absalom, Absalom* (43 wps) , *Intruder in the Dust* (39wps), and *Go Down, Moses* (20wps). The only notable exception is *Light in August* (15wps), which has Joe Christmas’s racial ambiguity as its main theme, but does not, on the whole complement this stylistically with long sentences as expected. Another confounding feature is that the other set of texts with above average sentence lengths is the Snopes trilogy (*The Hamlet*, *The Town*, *The Mansion*), which is actually notable for being almost entirely devoid of non-White characters.[[9]](#footnote-33) This discrepancy also underscores the limitations of this analysis. After all, just because a text features more non-White characters than other texts does not mean that they are central to the action, or appear in those long sentences.

### Character Types

In the Digital Yoknapatawpha database, the events data table stores the coincidence of characters and locations, where each event represent one action by a single or group of characters in one location. As the definition of an event is semantic and not syntactic, it effectively means that events are not coextensive with sentences. Occasionally, an event can run for several pages capturing multiple sentences or it can be only a few words in the middle of a sentence. Although technically involved, events can be reconciled with the sentences they appear in and the sentences in which they appear. This can be joined with the character table to indicate, which characters are present and mentioned in every sentence.[[10]](#footnote-35) This gives an approximation of the relationship between characters and sentences. Notably, there is no way of knowing what role each character plays in the event: that data simply does not exist. It could be that a character is simply in the background and is not the main part of the action. The baseline assumption is that a frequent recurrence of a character type indicates some salience even if they are not the focus of every event.

Since there are roughly ninety-thousand sentences in the entire corpus that can be anywhere from one to six thousand words long there is a really broad range of sentence types in which characters can appear. A Pearson correlation was used to establish the relationship between the types of characters present in a sentence and sentence length.[[11]](#footnote-36) The results are at odds with those of the text-level correlation in the previous section. Whereas the presence of characters with multiple ancestries tended to be predictive of sentence length at the level of the work, at the level of the sentence there is a strong correlation between sentence length and whether an upper class White man is present (0.86) or mentioned (0.742). As upper class White men tend to be the protagonists of these novels, it makes sense that appear in the longest sentences. What is quite telling is who is less likely to be present. It is true for both upper class White women, lower class White men, and free Black men that their correlation coeffiecent is higher when they are mentioned (0.625, 0.497, and 0.403 respectively) than when they are present (0.478, 0.311, and 0.147 respectively). The difference between these correlations is notable, as the strength of the correlation reduces by about 15%.[[12]](#footnote-37) That said, this drop does not mean that White women, lower class White men, and free Black men are necessarily being spoken about in these sentences. Mentioned in the database is a catch-all term for any character who is mentioned in an event but does not actually appear. It could be that the event only references a character once or that the entire event is about that character. Furthermore, correlation is not causation. If one of these characters is present it does not mean that the sentence length will get shorter, only that they are more likely to be present than present in longer sentences.

All those caveats aside, the correlations highlight a common scenario in Faulkner: two White men, usually upper class, discussing another person who is either not male, upper class, White or or all three. This includes Quentin and Mr. Compson talking about Caddy in *The Sound and the Fury*, and Ratliff and Gavin Stevens talking about the lower class Flem Snopes in *The* *Town* and *The* *Mansion*, and any number of White men talking about Black men. What falls away from this picture is the correlation between sentence length and any troubling of the racial binary of the Jim Crow South. There is only a weak correlation between mixed ancestry characters and sentence length. A plausible explanation for this is that the correlation between sentence length and character type was run across the entire corpus. Since, characters of mixed ancestry do not appear in all texts, their sentence length is effectively 0 for quite a number of texts. Accordingly, their individual presence in long sentences gets lost in the general mass of writing.

It is possible to tease out this relationship by performing a more fine-grained analysis and looking at character type and sentence length correlations within each work instead of across the corpus. The pattern that occurs here is quite telling. Essentially, upper class White men consistently have a moderate to strong correlation with longer sentences, but there are also other character types whose correlation suddenly becomes stronger in relation to sentence length. Thus, in the short story ‘Lion’ both upper class white men and the lone lower class male with Native American and White ancestry, Boon Hogganbeck, strongly correlate with longer sentences (0.152, 0.238). Still, this is probably because Boon happens to appear frequently in scenes with an upper class white male, Quentin Compson. This is why the average sentence length when an upper class White male is alone is 55 words-per-sentence, which drops to 21 when there are other characters present.[[13]](#footnote-38)

### Individual Characters

The final and most fine-grained way to consider sentence length is by individual characters. This shows which specific characters tend to be present or mentioned in lengthy sentences. As the previous data already indicates, the types of characters who are consistently present or mentioned in long sentences is quite limited. It should therefore be little surprise that the character who has the strongest correlation with sentence length is Quentin Compson (0.593), who throughout the *Sound and the Fury* and *Absalom, Absalom* is at the centre of the highly experimental narrative. Notably, Mr. Jason Compson, his father (0.476), also has a moderate correlation with sentence length when he is mentioned. The first woman on the list is Caddy Compson (0.446), and the first person white person who is not upper class is the murderer Vinson Gowrie (0.389) whose mention correlates with longer sentences. Also included is the mixed ancestry Charles Bon (0.385) who, importantly, may be present in the event, but is constructed by a White narrator, mainly Quentin.

## Conclusion

This renders an odd picture of sentence length and its attendant mid-sentence experimentation in Faulkner. On the surface, long sentence seem to be a hallmark feature of his writing, but texts with long sentences are not evenly distributed throughout the corpus, and within texts with long sentences they are the exception not the rule. While Faulkner’s own theory of writing suggests that the long sentence, the site of experimentation, is a generalized version of the human condition, it is a human condition that tends to be reserved for upper class White male characters and whomever happens to be present around them. It is therefore paradoxically at one and the same time a radical disavowal of temporality that unsettles a stable identity, but also a re-entrenchment of the very hierarchies that the disavowal aims to abjure. It is an experimentation that is at one and the same time radical and conservative.

Adams, Richard P. *Faulkner: myth and motion*. Princeton University press, 1968.

Aiken, Conrad. *William Faulkner: The Novel as Form*. Edited by Dean Morgan Schmitter, Mac Graw-Hill Book Company, 1973, pp. 45–52.

Attridge, Derek. “Molly’s Flow: The Writing of "Penelope" and the Question of Women’s Language.” *Modern Fiction Studies*, vol. 35, no. 3, 1989, pp. 543–65, <https://doi.org/10.1353/mfs.0.0210>.

Beck, Warren. *Faulkner: Essays*. University of Wisconsin Press, 1976.

Burgers, Johannes. *Absalom, Absalom! And the Digital Humanities: Charting Characters, Events, and Language*. Edited by Susan Scott Parrish, First edition, W.W. Norton & Company, Inc, 2023, pp. 677–92.

Chen, Peter Y., and Paula M. Popovich. *Correlation: Parametric and Nonparametric Measures*. Sage, 2002.

Cohen, Jacob. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed, L. Erlbaum Associates, 1988.

Corrigan, John, and John Padgett. *Faulkner’s ’the Bear’*. 2015, <https://faulkner.iath.virginia.edu/?text=BAR>.

DeVoto, Bernard. *"Witchcraft in Mississippi." Saturday Review of Literature, 15 (October 31, 1936), 3-4, 14*. Edited by M. Thomas Inge, Cambridge University Press, 1995, pp. 144–49.

Fanzun, Sebastian. *Faulkner and Formalism*. Edited by Sarah Gleeson-White and Pardis Dabashi, 1st ed., Cambridge University Press, 2022, pp. 19–35, <https://doi.org/10.1017/9781108881654>.

Fargnoli, A. Nicholas, and Michael Golay. *Critical Companion to William Faulkner*. Facts On File, Incorporated, 2009.

Faulkner, William. *A Fable*. 1st Vintage International ed, Vintage International/Vintage Books, 2011.

---. *Absalom, Absalom!* Knopf Doubleday Publishing Group, 2011.

---. *Absalom, Absalom!* Knopf Doubleday Publishing Group, 2011.

---. *English Department Faculty and Wives*. 13 May 1957, <https://faulkner.lib.virginia.edu/display/wfaudio13.html>.

---. *Intruder in the Dust*. Vintage International, 2011.

---. *Local and UVA Communities, Tape 1*. 30 May 1957, <https://faulkner.lib.virginia.edu/display/wfaudio18_1.html>.

---. *"Miss Zilphia Gant," Bound Typescript 18 p.*

---. *The Hamlet: The Corrected Text*. 1st Vintage International ed, Vintage Books, 1991.

---. *The Sound and the Fury: The Corrected Text*. 1st Vintage International ed, Vintage Books, 1990.

---. *Uncollected Stories of William Faulkner*. Edited by Joseph Blotner, Vintage International edition, Vintage International, 1997.

---. *Uncollected Stories of William Faulkner*. Edited by Joseph Blotner, Vintage International edition, Vintage International, 1997.

---. *Undergraduate Literature Class*. 3 Nov. 1957, <https://faulkner.lib.virginia.edu/display/wfaudio04.html>.

---. *Virginia Colleges Conference, Tape 1*. 15 Apr. 1957, <https://faulkner.lib.virginia.edu/display/wfaudio07_1.html>.

---. *Virginia Colleges Conference, Tape 2*. 15 Apr. 1957, <https://faulkner.lib.virginia.edu/display/wfaudio07_2.html>.

Gagolewski, Marek. “Stringi: Fast and Portable Character String Processing in R.” *Journal of Statistical Software*, vol. 103, no. 2, 2022, <https://doi.org/10.18637/jss.v103.i02>.

Hannon, Charles. “The Function of Function Words in William Faulkner’s as "i Lay Dying".” *Faulkner Journal*, vol. 29, no. 1, 2015, pp. 3–21, <https://doi.org/10.1353/fau.2015.0006>.

Hlavsa, Virginia V. James. *Faulkner and the Thoroughly Modern Novel*. University Press of Virginia, 1991.

Hubert, Mia, and Ellen Vandervieren. “An Adjusted Boxplot for Skewed Distributions.” *Computational Statistics & Data Analysis*, vol. 52, no. 12, 2008, p. 51865201, <https://doi.org/10.1016/j.csda.2007.11.008>.

Ihrmark, Daniel, and Johan Nilsson. “A Corpus Stylistic Analysis of Development in Hemingway’s Literary Production.” *The Hemingway Review*, vol. 40, no. 2, 2021, pp. 71–93, <https://doi.org/10.1353/hem.2021.0005>.

Lockyer, Judith. *Ordered by words: language and narration in the novels of William Faulkner*. Southern Illinois Univ. Press, 1991.

Meyerson, Gregory, and Jim Neilson. “Pulp Fiction: The Aesthetics of Anti-Radicalism in William Faulkner’s "Light in August".” *Science & Society*, vol. 72, no. 1, 2008, pp. 11–42, <https://doi.org/10.1521/siso.2007.72.1.11>.

Millard, Steven P. *EnvStats: An r Package for Environmental Statistics*. Springer, 2013.

Newman, Mej. “Power Laws, Pareto Distributions and Zipf’s Law.” *Contemporary Physics*, vol. 46, no. 5, Sept. 2005, pp. 323–51, <https://doi.org/10.1080/00107510500052444>.

Nunberg, Geoffrey. *The Linguistics of Punctuation*. Center for the Study of Language (CSLI), 1990.

---. *The Linguistics of Punctuation*. Center for the Study of Language (CSLI), 1990.

Olsen, Kathryn. “Raveling Out Like a Looping String: As i Lay Dying and Regenerative Language.” *Journal of Modern Literature*, vol. 33, no. 4, 2010, pp. 95–111, <https://doi.org/10.2979/jml.2010.33.4.95>.

Parker, Robert Dale. “Red Slippers and Cottonmouth Moccasins: White Anxieties in Faulkner’s Indian Stories.” *Faulkner Journal*, vol. 18, no. 1/2, 2002, pp. 81–99, <https://www.jstor.org/stable/24908269>.

Polk, Noel. *Children of the Dark House: Text and Context in Faulkner*. Univ. Press of Mississippi, 1998.

Railton, Stephen. *Manuscripts &c: “Miss Zilphia Gant"*. 2017, <https://faulkner.drupal.shanti.virginia.edu/node/18509?canvas>.

Rudnicka, Karolina. *Variation of Sentence Length Across Time and Genre: Influence on Syntactic Usage in English*. Edited by Richard J. Whitt, John Benjamins Publishing Company, 2018.

Schou, Karsten. “The Syntactic Status of English Punctuation.” *English Studies*, vol. 88, no. 2, Apr. 2007, pp. 195–216, <https://doi.org/10.1080/00138380601042790>.

Scott, Arthur L. “The Faulknerian Sentence.” *Prairie Schooner*, vol. 27, no. 1, 1953, pp. 91–98, <https://www.jstor.org/stable/40624529>.

Sun, Kun, and Rong Wang. “Frequency Distributions of Punctuation Marks in English: Evidence from Large-Scale Corpora.” *English Today*, Dec. 2018, <https://doi.org/10.1017/S0266078418000512>.

Thomas, Sara Gabler. “Queer Formalism: Synesthetic Storytelling in Monique Truong and William Faulkner.” *Faulkner Journal*, vol. 30, no. 1, 2016, pp. 39–61, <https://doi.org/10.1353/fau.2016.0010>.

Van O’Connor, William. “Faulkner.” *The Georgia Review*, vol. 12, no. 1, 1958, pp. 83–86, <https://www.jstor.org/stable/41395494>.

Weinstein, Philip M. *Faulkner’s subject: a cosmos no one owns*. Cambridge university press, 1992.

1. Before it became possible to do computational analysis of texts, there have been any number of precursors that anticipate the methodology. The most extensive of these are the textual concordances published by the Faulkner Concordance Advisory Board. The work of Virginia Hlavsa in 1991 also uses word counts to make its argument (Hlavsa 51–53). More recently, Charles Hannon has used these techniques with *As I Lay Dying* (Hannon) [↑](#footnote-ref-20)
2. In an earlier piece on punctuation in *Absalom, Absalom!* sentence length was calculated by accounting for embedded quotation. This created sentences of 49 words-per-sentence, which is substantially higher than the 43 words-per-sentence arrived at in this paper that counts embedded quotes as individual sentences (Burgers 688). [↑](#footnote-ref-24)
3. There is an extensive critical literature on the “flow” of Molly’s soliloquy exhaustively documented by Derek Attridge in “Molly’s Flow: The Writing of”Penelope” and the Question of Women’s Language”. [↑](#footnote-ref-26)
4. As previously indicated, *Requiem for a Nun* was removed from this data set, because it was too much at variance with the other texts. [↑](#footnote-ref-27)
5. In theory, this rule does not necessarily to specific punctuation marks that are more strictly governed by the rules of grammar, such as colons, which generally do not exceed one per sentence. [↑](#footnote-ref-28)
6. Deciding that a data point is an outlier is an ongoing debate among mathematicians (Hawkins 9-12), and, ultimately, the decision to say that a data point is outside the expected range is still intepretive. That said, two different outlier tests were were run: Rosner’s multiple outlier test (Millard 1106) and and the Adjusted Box plot method (Hubert and Vandervieren). All of the calculations and results for these tests are available here: [↑](#footnote-ref-29)
7. Concatenating every possibility of character produces a very large number of types. As some of these, such as Asian men, have a very low occurrence they were eliminated from analysis. Similarly, “group” variables were also eliminated because they are non-countable. Their presence may indicate a few people to the entire town, and running analysis on these would be misleading because it is not always a like with like comparison. [↑](#footnote-ref-31)
8. Pearson correlations produce two important numbers significance () and the correlation coefficient (). Significance indicates the percent chance that the relationship is not random. Unless otherwise indicated the threshold for significance was set at <.05. The magnitude of a Pearson correlation runs from -1 to 1, where a negative correlation indicates an inverse relationship and a positive correlation indicates a direct relationship. Though interpreting correlation varies substantially from discipline to discipline (Chen and Popovich 12), the most commonly agreed upon ranges are roughly a correlation between .3 and .5 or between -.3 and -.5 is considered a medium effect, and above .5 or below -.5 is considered a large effect (Cohen 79–80). [↑](#footnote-ref-32)
9. Only 6% are not-White compared to 22% in the other texts. [↑](#footnote-ref-33)
10. The *DY* database stores the first 8-10 words of each event. This makes it possible to delimit where an event begins and ends in each part of a digital version of the text. Occassionally, this requires manual realignment due to errors in the database or errors in the text. In either case, these realignments are statistically insignificant for the overall picture of punctuation patterns. [↑](#footnote-ref-35)
11. This is a very large sample size, and the -value, the odds that the correlation is random, tends to be quite low. [↑](#footnote-ref-36)
12. Correlation coefficients are not linear in strength. Thus, a coefficient of .4 is not twice as strong as on of .2. The coefficient of determination, , explains the variance between the two relationship between the two variables (Chen and Popovich 12). For these numbers, the average difference in the values is 15%. [↑](#footnote-ref-37)
13. It is not possible to run a correlation test on only those sentences where a character is alone, because there are too few data points. Generally, though it is true that a character alone, irrespective of race, class, or gender tends to be embedded in longer sentences than in a group. That said, the plurality of events where a character was alone featured upper class white males (33%) [↑](#footnote-ref-38)