

The road to hell is paved with adverbs.

—STEPHEN KING

n literary lore, one of the best stories of all time is a mere six words. "For sale: baby shoes, never worn." It's the ultimate example of *less is more*, and you'll often find it attributed to Ernest Hemingway.

It's unclear whether it was in fact Hemingway who penned these words—the story of its creation did not appear until 1991—but it's natural that writers and readers would want to attribute the story to the Nobel winner. He's known for his economical prose, and the shortest-of-short stories is, at the very least, emblematic of his style.

Hemingway's simple style was an intentional choice. He once wrote in a letter to his editor, "It wasn't by accident that the Gettysburg address was so short. The laws of prose writing are as immutable as those of flight, of mathematics, of physics." He believed that writing should be cut down to the bare essentials and that extra words end up hurting the final product.

Ernest Hemingway is far from alone in this belief. The same idea is raised in high-school classrooms and writing guides of

every variety. And if there's one part of speech that's the worst offender of all, as anyone who's ever had an exacting English teacher will know, it's *the adverb*.

After listening to enough experts and admirers, it's easy to come away with the impression that Hemingway is the paragon of concision. But is this because he succeeded where others were tempted by extraneous language, or is he coasting on reputation alone? Where does Hemingway rank, for instance, in his use of the dreaded adverb?

I wanted to find out if he lived up to the hype. And if not, who does use the fewest adverbs? Which author uses them the most? Moreover, when we look at the big picture, can we find out whether great writing does indeed hew to those efficient "laws of prose writing"? Do the best books use fewer adverbs?

I looked around and found that no one had ever attempted to determine the numbers behind these questions. So I sought to find some answers—and I started by analyzing the almost one million words in Hemingway's ten published novels.

If Hemingway believes that the "laws of prose writing are as immutable as those of flight, of mathematics, of physics," then I'd like to think he'd find this mathematical analysis equal parts illuminating and outlandish.

It's outlandish at first glance because of the way we study writing. Many of us have spent days in middle school, high school, and college English classrooms dissecting a single striking excerpt from a Hemingway novel. If you want to study a great author's writing, their most remembered passages are often the best place to start. Looking at a spreadsheet of adverb frequencies, on the other hand, won't teach you much in the way of writing a novel like Hemingway.

But from a statistician's point of view, it's just as outlandish to focus on a small sample and never look at the whole picture. When you study the population of the United States, you wouldn't look

at just the population of a small town in New Hampshire for an understanding of the entire country, no matter how emblematic of the American spirit it may seem. If you want to know how Hemingway writes, you also need to understand the words he chooses that have not been put under the microscope. By looking at adverb rates throughout all his books, we can get a better sense of how he used language.

So instead of digging through snippets of Hemingway's text and debating specific spots where he chose to use or shirk adverbs, I used a set of functions called Natural Language Toolkit to count the number of adverbs in all of his novels. The toolkit relies on specific words and the relationships between them to tag words with a part of speech. For example, here's how it processes the previous sentence:

The toolkit relies on specific words and the relationships

between them to tag words with a part of speech.

It's not 100% perfect—so all the numbers below should be seen with that wrinkle in mind—but it's been trained on millions of human-analyzed texts and fares as well as any person could be expected to do. It's considered the gold standard in sussing out if a word is an adjective, adverb, personal pronoun, or any other part of speech.

So what do we find when we apply the toolkit to Hemingway's complete works?

In all of Hemingway's novels, he wrote just over 865,000 words and used 50,200 adverbs, putting his adverb use at about 5.8% of

all words. On average, for every 17 words Hemingway wrote, one of them was an adverb.

This number without context has no meaning. Is 5.8% a lot or a little? Stephen King, an outspoken critic of adverbs, has a usage rate of 5.5%.

It turns out that by this standard King and Hemingway are not leaps and bounds ahead of other writers. Looking at a handful of contemporary authors who one might assume (based on stereotype alone) would use an abundance of adverbs, we see that King and Hemingway are not anomalous. E L James, author of the erotica novel *Fifty Shades of Grey*, used adverbs at a rate of 4.8%. Stephenie Meyer, whom King has called "not very good," used adverbs at a rate of 5.7% in her Twilight books, putting her right between the horror master and the legendary Hemingway.

Expanding our search, Hemingway used more adverbs than authors John Steinbeck and Kurt Vonnegut. He used more adverbs than children's authors Roald Dahl and R.L. Stine. And, yes, the master of simple prose used more adverbs than Stephenie Meyer and E L James.

All the sentences above are true—but they also need a giant asterisk next to them and a full explanation. Because the answer is not as simple as the numbers above first suggest.

Those tallies are counts of total adverb usage. An adverb is any word that modifies a verb, adjective, or another adverb—and no adverbs were excluded or excused. But when Stephen King says, "The adverb is not your friend," he's not talking about any word that modifies a verb, adjective, or another adverb. In the sentence "The adverb is not your friend," the word not is an adverb. But not is not King's issue. Nobody reads "For sale: baby shoes, never worn" and thinks never is an adverb that should have been nixed.

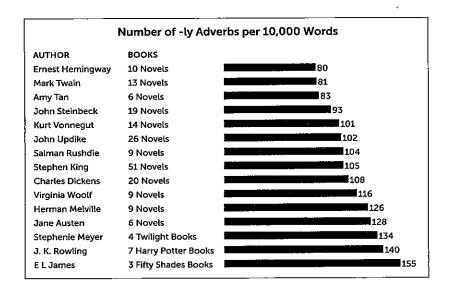
When King rails against adverbs in his book *On Writing*, he describes them as "the ones that usually end in -ly." From a statistical standpoint his "usually" isn't quite true (depending on the

author, around 10 to 30% of all adverbs are ones that end in -ly) but it is true that the adverbs ending in -ly are the ones that tend to stick out.

Chuck Palahniuk, best known as the author of *Fight Club*, has written against -ly adverbs as well. When discussing the importance of minimalism in his book *Stranger than Fiction*, Palahniuk writes, "No silly adverbs like 'sleepily,' 'irritably,' 'sadly,' please." His general argument is that writing should allow us to know when a character is sleepy, or irritable or sad, by using a broader set of clues rather than a single word. Using -ly adverbs goes too far, telling the reader what they should think instead of setting up the scene so that the meaning becomes clear in context.

By narrowing our search to just -ly adverbs, we can cut to the heart of the debate. And when we do, the picture flips. For every 10,000 words E L James writes, 155 are -ly adverbs. For Meyer the count is 134, while King averages 105. And Hemingway, living up to his reputation, comes in at a scant 80.

Below, for the sake of comparison, is a breakdown of adverbuse among 15 different authors.



Looking at this strict definition of the "bad" kind of adverb, Hemingway indeed comes out as one of the greats. As we continue to explore in this chapter, whenever I use the term *adverb*, I will be referring to this "bad" sort—the -ly adverb.

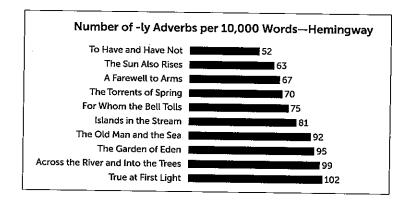
## Was Hemingway Right?

The list on the previous page includes a variety of writers, from Nobel Prize winners to viral bestsellers. Hemingway may emerge as a titan of unadorned prose, just as the common perception of him would suggest. But any broader pattern is not so clear. E L James lands at the top of the scale, but greats like Melville and Austen also clock in toward the higher end. By adding more data points, would we be able to pinpoint a reliable pattern in adverb usage?

I wanted to find out whether an author's adverb rate reflects anything more than just personal style or preference. I was curious: Could Hemingway have been right about the "laws of prose"? Is there any meaningful relationship between the quality of a book and how often it uses adverbs?

To start answering these questions, it's important to note that just as different authors vary in their use of adverbs, so do different books by the same author. The rate of these -ly adverbs is rare—under 2%—even for authors who use them more than other scribes. And there is often great variation from book to book within an author's career.

For instance, looking at Hemingway's novels, we see a wide range. Several of his books have adverb rates much lower than most authors ever come close to approaching, while other books bounce around the average usage rate of other authors. *True at First Light*, a novel about Hemingway's experiences in Africa, is his novel with the highest adverb usage—and it's one released thirty years after his death.



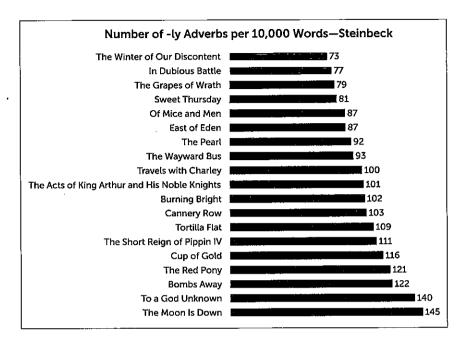
True at First Light was received by critics with negative reviews. It was unfinished at the time of Hemingway's death and edited into shape by his son. Some saw its publication as an unnecessary addition to the canon. Is it a coincidence that it is also his work with the most adverbs?

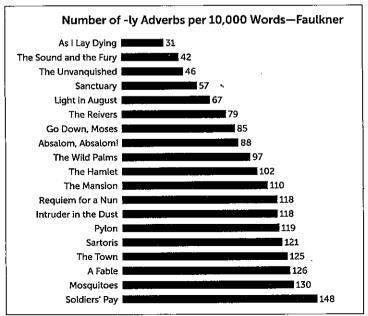
It's of course a poor criterion to judge a book on nothing but its adverb rate, but looking at Hemingway's complete works, we see that most of his classics are also the texts in which he uses the fewest adverbs. The Sun Also Rises, A Farewell to Arms, and For Whom the Bell Tolls have some of the lowest rates—and are considered to be among Hemingway's best. The Old Man and the Sea, which won the author the Pulitzer Prize and is often named Hemingway's best work, is the exception.

The two American authors to win the Nobel Prize in literature within a decade of Hemingway are William Faulkner and John Steinbeck, and we can pick apart their stats as well.

For Steinbeck, the rate of adverb usage again matches up well to perceptions of his work. *The Grapes of Wrath*, perhaps his most popular work, places third on the list. *Of Mice and Men* and *East of Eden* also land toward the low end.

For Faulkner, the pattern is again present. His most celebrated work, *The Sound and the Fury*, ranks second with a low 42 ad-





verbs per 10,000. As I Lay Dying and Light in August also come in at the top, while Absalom, Absalom! is just under his average as well.

But this is just three authors. How far does the pattern go? If we expand outward, do the best books by the best authors use fewer adverbs on average?

The authors selected for the tables on the previous pages, and the books highlighted, are some of my own favorites, chosen based on my own preferences. To test whether adverb rates have any correlation with writing quality, I would need a larger set of books and writers—and I would need them to be considered "great" by a consensus of readers.

To build a new sample, I turned to four different lists of the best twentieth-century literature: the *Library Journal* list, the Koen Book Distributor's list, the Modern Library List, and the Radcliffe Publishing Course list. All four lists rank at least 100 works of fiction in English literature. These four lists were also used by Stanford librarian Brian Kunde in his attempt to quantify the best book of the twentieth century (by his scoring it's *The Great Gatsby*). For my purposes, if a book was included on at least two of the four lists I deemed it a consensus "great" book. I then selected the authors who had at least two of these, so I would be able to compare their "great" books to the "non-great." (Of course, this method leaves out a lot of excellent authors, but I needed something approaching "objective" and I needed authors with multiple works.)

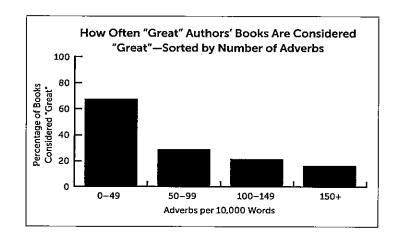
The result is 15 consensus "great" authors with a total bibliography\* of 167 fiction books, of which 37 books were considered "great" by virtue of being on multiple top-100 lists. Here, I've listed those 37 "great books."

<sup>\*</sup> Some of Sinclair Lewis's novels could not be found in digital form and were excluded.

The Consensus "Great" Books		
Willa Cather Death Comes for the Archbisho My Ántonia	E. M. Forster op A Passage to India A Room with a View Howards End	Toni Morrison Beloved Song of Solomon
Joseph Conrad Heart of Darkness Lord Jim	Ernest Hemingway A Farewell to Arms For Whom the Bell Tolls The Old Man and the Sea The Sun Also Rises	George Orwell Animal Farm 1984
Theodore Dreiser An American Tragedy Sister Carrie	James Joyce A Portrait of the Artist as a Young Man Ulysses	Ayn Rand Atlas Shrugged The Fountainhead
William Faulkner As I Lay Dying Light in August The Sound and the Fury	Finnegans Wake  D. H. Lawrence  Lady Chatterley's Lover  Sons and Lovers	John Steinbeck Of Mice and Men The Grapes of Wrath
F. Scott Fitzgerald Tender Is the Night The Great Gatsby	Women in Love Sinclair Lewis Babbitt Main Street	Edith Wharton Ethan Frome The Age of Innocence The House of Mirth

The test I conducted was simple but revealing, aimed at parsing whether there is a noticeable difference between the best books and the rest. I combined all 167 books written by the consensus "great" authors, and I broke them down into groupings of 50 adverbs per 10,000 words. I then looked at how many books in each group were selected as "great" versus "nongreat." The graph on the following page charts the results. Books with 0–49 adverbs per 10,000 words were considered great by critics 67% of the time. Those in the 50–100 range were selected as great 29% of the time. On the far end, those with 150-plus received the honor just 16% of the time.

The downward slope of the graph backs up the advice of Hemingway, King, and countless other writers. While far from absolute, there's a clear trend as adverb use increases. The best books—the



greats of the greats—do use a lower rate of -ly adverbs. On the other hand, an overuse of adverbs has resulted in "great" books at a far scarcer frequency.

## Author by Author, Adverb by Adverb

Looking at the best books all in one graph gives us a compelling picture, but it is still just part of the picture. The greats may tend to use fewer adverbs, but it's also clear that a book or author doesn't *need* to adopt the same low rate to be great. Consider Sinclair Lewis, a Nobel Prize winner and one of our consensus "great" authors, who writes at an average adverb rate of 142 per 10,000 words. That's *a lot*—75% more than Hemingway averaged.

At the very least, Lewis is an outlier—perhaps even an argument against any general trend. But when we dig into his work, we find a pattern that seems to apply even to the previous section's outliers. What's interesting about Lewis's work is that his two best books—*Main Street* and *Babbitt*, his two consensus "great" books—use fewer adverbs than any of his other novels. In other words, even though Lewis uses a very high rate of adverbs, his most popular writing is his most concise.

With Hemingway, Steinbeck, and Faulkner, we have already seen that their masterpieces also tend to be their books with fewer adverbs. And, looking beyond this trio, we find plenty of similar examples throughout the great authors:

- The Great Gatsby is F. Scott Fitzgerald's book with the lowest adverb rate. His second most popular novel, Tender Is the Night, is his book with the second lowest rate.
- Toni Morrison's most acclaimed novel, Beloved, is tied as her book with the fewest adverbs.
- A Tale of Two Cities and Great Expectations beat out the other 13 Charles Dickens novels to have the lowest and second lowest adverb rates.
- Kurt Vonnegut wrote 14 novels, and his three most acclaimed are Cat's Cradle, Slaughterhouse-Five, and Breakfast of Champions. They rank first, second, and third in least adverb usage out of all his works.
- John Updike authored 26 novels. The four novels with the smallest adverb rate were all four books in his Pulitzer Prize winning Rabbit tetralogy.

The string of examples goes on, but there are also notable exceptions. D. H. Lawrence, for instance, wrote two "great books," in *Lady Chatterley's Lover* and *Women in Love*, that use *more* adverbs than any of his other works. If we continue to search, we can find anecdotal evidence for either side.

The Sinclair Lewis question, then, was the next big trend I wanted to investigate. Lewis, it seemed, might reveal an even broader truth about how authors use and abuse their adverbs: Regardless of an author's fondness for adverbs—whether their natural rate skews high or low—are they at their most successful when they're most concise?

To test this, I would need to go beyond the simple distinction between "great" and "non-great" books. I would need to be able to compare all novels within an author's bibliography on a sliding scale, measuring how good the "great" books are and how bad the "nongreat" are. Doing so, it would be possible to chart out whether—book by book, within an individual author's career—there's a broader correlation between adverb use and writing quality.

How, though, can you compare any two books in an objective manner? When neither appears on any critic's "best of" list, how do we know with any reliability whether Steinbeck's *The Pearl* can be considered better than his novel *To a God Unknown*?

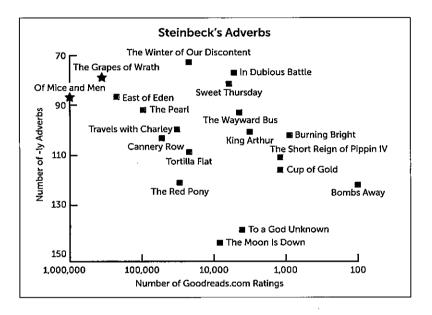
The solution I settled upon was to dig into the ratings found on book reviewing sites like Amazon or Goodreads. Goodreads .com is a website where people go to rate, discuss, and catalog books. Here, a popular book can have more than one million ratings, a quantity much larger than the same book would receive on Amazon. Because of its size, we'll explore Goodreads ratings.

In particular, I've chosen to focus on how *many* Goodreads ratings a book has. It's not a perfect metric, but it gives a relative sense of the reception and popularity of a book. In fact, it's a better metric than the book's average numerical rating—the average number of stars it receives from reviewers.\* Using Goodreads data, we're no longer stuck with the binary of a book being "great" or "not great." We can instead get a sense of a book's popularity on a spectrum, which gives us a much fuller picture of its quality and status.

We can then go back to our Steinbeck and Faulkner graphs and bulk them up with more depth. The books farthest to the left have the most ratings while the books to the right have the least. The books closest to the top have the fewest adverbs while those

<sup>\*</sup> It's more wonky than we can get into here, but for a detailed explanation of why average rating falls short, head to the Notes section on p. 251.

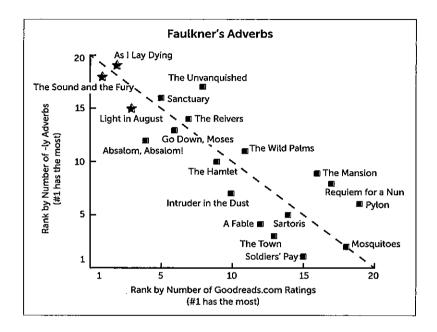
at the bottom have the most. For Steinbeck, *The Grapes of Wrath* has many ratings and few adverbs so it's in the upper left. *Bombs Away* has few ratings but lots of adverbs, so it's off in the lower right-hand corner.



If the correlation were perfect, the books would form a pattern always trending down and to the right when looking at the chart. It's not exact, but the correlation exists.

The number of ratings in the graph above is displayed using a logarithmic scale, meaning a book with 100,000 ratings will be just as far off from a book with 10,000 ratings as a book with 10,000 ratings would be from one with 1,000. Without the logarithmic scale, the books would be too far apart to make any sense of it. If you're having trouble wrapping your head around the logarithmic scale, the Faulkner graph shows the same trend but instead breaks the data down into rankings—the book with the most Goodreads ratings is considered number one on the horizontal axis, then the

next highest is number two, and so on. If the correlation were perfect, every book would fall on the dashed line.

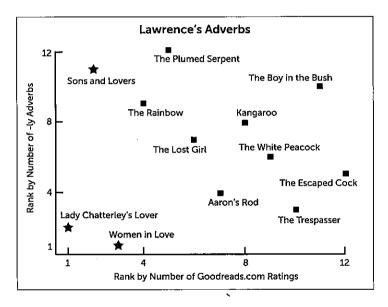


Even if you were to look at just the "non-great" books—excluding As I Lay Dying, The Sound and the Fury, and Light in August—the relationship holds in stunning fashion. The empty lower left-hand quadrant shows the complete lack of Faulkner books that have both high-ly adverb rates and a noteworthy reputation.

This amazing trend does not hold for every author. If there was one author poised to buck the trend, based on the complete listings of adverb use above, it would be D. H. Lawrence. The Englishman was unique among our 15 authors in that his book with the *most* adverb usage was considered "great." And his book with the second-most adverbs was considered "great" as well. You might then expect that the rest of his work would follow that same trend,

forming a line from the bottom left corner to the top right (the opposite of Faulkner's).

But this is not the case. While his 12 books are not enough to draw definite conclusions from, any clear pattern (for or against a relationship) is not seen in Lawrence's more jumbled chart.



Though Faulkner's graph may appear to have an undeniable pattern at first glance, the brain can sometimes play tricks searching for patterns in images, so it's always better to test for significance to see what the raw numbers say. In Faulkner's case the numbers back up the eye test. There is a correlation between adverb use and his book ratings (going back to log scale) when tested. Lawrence's graph, on the other hand, has no pattern to speak of. But which of these graphs is the norm? Is Lawrence the outlier, or are Faulkner and Steinbeck anomalies?

The bigger picture comes together when we combine all the authors into a single graph and normalize their adverb usage and

Goodreads popularity.\* With this approach we can look at all our great authors and all their books, and we can ask whether there is a correlation between adverb use and book quality within each author's career.

In the full sample of 167 books, we find that there is indeed a correlation. The pattern isn't perfect, but the connection is striking—and it goes well beyond the variation we could expect due to chance alone. Up and down the sample, we find that authors' books with the least adverbs have been their most popular, and their books with more adverbs have tended to earn lower ratings.

The chart below illustrates the general pattern, as well as highlighting the large number of outliers. The top left-hand corner shows books that ranked in the top half of an author's most popular works and are in the bottom half of adverb usage. Fifty books fall in this range. In contrast, just thirty-one books are among the most popular half but also high in adverbs. The hits are concise, while the wordier novels are often forgotten.

High Goodreads Ranking	Low Goodreads Ranking	
Low in Adverbs	Low in Adverbs	
50	30	
High Goodreads Ranking	Low Goodreads Ranking	
High in Adverbs	High in Adverbs	
31	50	

Note: The numbers do not sum to 167 books because some books were at the median—meaning they could not be categorized.

<sup>\*</sup> Unlike the previous section's aggregate graph, where books from different authors were combined unadjusted, the normalization here allows us to compare authors of different levels of popularity and adverb rate without any outliers skewing the combined chart. The authors are treated as if they have equal rates and popularities, so that we can concentrate on the trends within each author's work.

## Pros vs. Amateurs

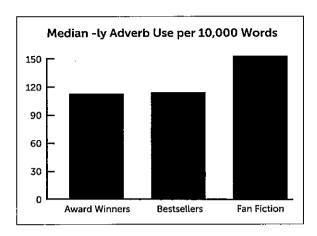
We've now seen that adverb use does play a role in the work of the canon's best authors and their greatest works. At the pinnacle of the literary world, the standout books indeed rely on fewer adverbs. And even within each author's own works, the books that use the least adverbs have been the most successful.

But there was one more question on my mind: What about the rest of us?

Before I'd be satisfied, I wanted to find out how great writers compare to the average writer when it comes to adverb use or abuse. Do Hemingway's "laws of prose" apply across the whole of the literary universe, from the award winners and bestsellers to the amateurs? I set up one final showdown to find out.

I downloaded more than 9,000 novel-length fan-fiction stories (of 60,000-plus words) from fanfiction.net. This would be my "amateur" group, consisting of all stories written between 2010 and 2014 in the 25 most popular book universes (ranging from Harry Potter to Twilight to *Phantom of the Opera* to Janet Evanovich's books). People writing stories this long are committed to their work, and many of them are strong writers. But on average, they're not at the level of the bestsellers or the award winners of the literary world. So I compared the fan-fiction sample to all of the books that have ranked number one on the *New York Times* bestseller list since 2000, and also to the 100 most recent winners of major literary awards.\*

When set side by side, the difference is clear. The median fan-fiction author used 154-ly adverbs per 10,000 words, which is much higher than either of the professional samples. The 300-plus megahits in the bestseller category averaged just 115-ly adverbs



per 10,000 words. And the 100 award winners have a median of 114-ly adverbs. It's not an apples-to-apples comparison, but the novels that sell well in bookstores come in with 25% fewer adverbs than the average novel that amateur writers post online. Less than 12% of all number one bestsellers had more than 154 adverbs, even though half of all fan fiction does.

The results of this chapter are one half common sense and one half mind-blowing.

Most writers and teachers will tell you that adverbs are bad. This is not a controversial stance to take. In many ways, the statistics presented above are just a confirmation of what we already knew.

But the fact that their use is somehow correlated with quality on a measurable level—even when just the best writers are being examined—is still shocking. It might not be a surprise that some beginner writers use adverbs as a crutch more often than professional writers, and that these traits may sometimes be noticeable. But even when looking at the life's work of the best writers, the effect is present.

<sup>\*</sup> The selections for these award-winning books is described in detail in Chapter 2.

A statistical correlation, of course, does not imply causation. Fitzgerald's *The Great Gatsby* used 128 adverbs per 10,000 words while his lesser-known *The Beautiful and Damned* used 176. If you picked up *The Great Gatsby* and stuck in 200 more adverbs, a bit less than one a page, it would have a higher rate than *The Beautiful and Damned*. Would this version of the book still be celebrated? What if you trimmed down the adverbs from *The Beautiful and Damned*? Would Leonardo DiCaprio be ready to suit up for the role of Anthony Patch?

The answer of course is that it's not so simple. Adverb rate alone could not have such a direct impact on the success of a book. There are thousands and thousands of other aspects of writing in play. The Hemingway adverb stereotype may be true, but there are notable counterexamples—authors who have written successful books when increasing their adverb usage. Nabokov's *Lolita*, for instance, has more adverbs than any of his other eight English novels.

One possible explanation for the overall trend we're seeing is that adverbs are an indicator of a writer's focus. An author writing with the clarity needed to describe vivid scenes and actions without adverbs, taking the time to whittle away the unnecessary words, might also be spending more time and effort making the rest of the text as perfect as possible. Or if one has a good editor, these words may be weeded out.

The "focus" hypothesis finds some support from the true master of writing without adverbs. And it's not Hemingway.

The numbers revealed an overlooked champion. Combing through a large number of authors, there was but one author on the list of "greats" who outdid Hemingway: Toni Morrison. She may be a Nobel and Pulitzer Prize winner just like Hemingway, but her place at the height of concise writing isn't often cited in English classrooms. Her adverb rate of 76 edges out Hemingway's 80, and puts her

wall ahaad of athera lile Stainhads Duahdia Salinger and Wharton

Morrison has said in multiple interviews that she doesn't use adverbs. Why? Because when she's writing at her best, she can do without: "I never say 'She says softly," Morrison tells us. "If it's not already soft, you know, I have to leave a lot of space around it so a reader can hear that it's soft."

There you have it. And while I have no hard evidence that the logic of adverb usage carries over to wacky statistics-based prose, I went through the text of this chapter to search for -ly adverbs after writing 5,000 words on how awful they were. I found that in most cases they were unneeded. They often blunted the impact of my sentences. I deleted all -ly adverbs that were not used when quoting or citing others.

As a result, if you excuse the ones in quotes, you will find no -ly adverbs in this chapter. This makes for a usage rate of 0 per 10,000 that would rank this text ahead of (or tied) with all other texts ever written. Does that make this chapter, regardless of content, a step above average? Here we've found the limits of our statistics. But when trying to write standout prose, it can't hurt to deliberately avoid the troublesome part of speech.