ARTS | BOOKS | BOOKSHELF

What Did Tolkien Have Against the Word 'She'?

The writer uses the word 'he' 1,900 times in 'The Hobbit.' The word 'she' appears just once. Jeff Baker reviews "Nabokov's Favorite Word Is Mauve" by Ben Blatt.

J.R.R. Tolkien in 1955. *PHOTO: GETTY IMAGES*

By **JEFF BAKER**
March 14, 2017 6:16 p.m. ET

When the Federalist Papers were published in 1788, the author was listed simply as Publius. But the 85 essays defending the new Constitution, as Americans found out after Alexander Hamilton's death in 1804, were in fact written by Hamilton, James Madison and John Jay. Hamilton wrote the overwhelming majority, as he made sure everyone knew in a note written shortly before his fateful duel with Aaron Burr. Hamilton claimed credit for 63 essays; he said Jay penned five and Madison wrote 17.

But when Madison later said that he had actually written 12 of the essays that Hamilton claimed credit for, the authorship of each of the Federalist Papers became a subject of intense debate. Historians argued about the disputed essays for some 150 years until two statisticians, Frederick Mosteller and David Wallace, used word frequency and probability to prove that all 12 essays in question were written by Madison. The 1963 breakthrough came when the statisticians realized—after months of work counting words by hand—that Hamilton used "while" but never "whilst" and Madison did the opposite. Dozens of other word patterns confirmed Madison's authorship. The correct totals: Hamilton wrote 51 essays, Madison 29, and Jay five.

Messrs. Wallace and Mosteller's research became a foundation of what is called stylometry, the study of language based on recurring word patterns, expressions and punctuation. Such textual analysis has practical applications in law, politics, advertising and, perhaps most powerfully today, search engines. Every comma is one more tiny piece of Big Data that establishes an online identity as distinct as a fingerprint. Nowadays it would take a few minutes on a laptop to figure out who wrote the Federalist Papers.

In "Nabokov's Favorite Word Is Mauve: What the Numbers Reveal About the Classics,

Bestsellers, and Our Own Writing," Ben Blatt, a writer with a degree in applied mathematics, uses computer-assisted text analysis to find out, for example, how many adverbs are in all the novels written by Ernest Hemingway, Stephen King and Mark Twain. Data-driven journalism is all the rage, and Mr. Blatt's goal is to release the hounds of stylometry into the fiction stacks and see what happens. We know that D.H. Lawrence loved animal similes—"like a caged hawk," he writes in "The Rainbow"; "like an indignant turkey," he writes in "The Lost Girl"—but did he love them twice as much as Jack London and 10 times as much as Edith Wharton?

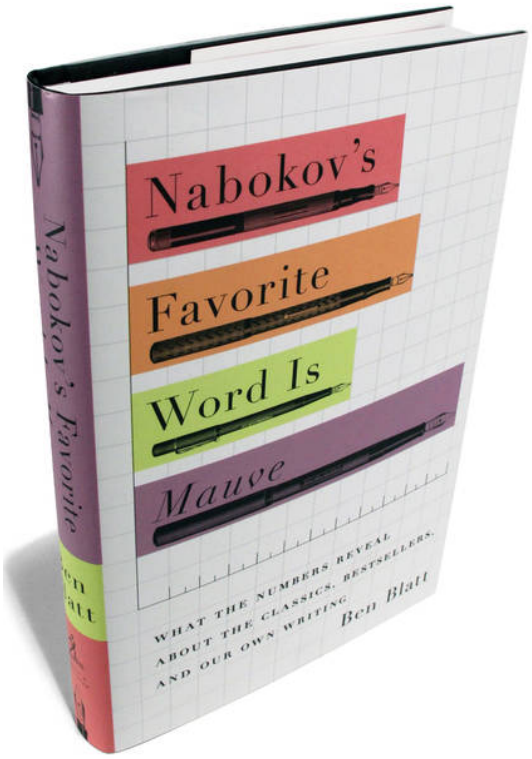

PHOTO: WSJ

NABOKOV'S FAVORITE WORD IS MAUVE

By Ben Blatt
Simon & Schuster, 271 pages, $25

He did, as Mr. Blatt proves in his fitfully amusing but frequently maddening study. Mr. Blatt is a diligent counter of adverbs and animal similes, but his idea of scientific rigor is to pose a question—Do English writers use "bloke," "blimey" and "brilliant" more often than American writers?—and attempt to answer it by searching for those words across thousands of books. Once he comes to a conclusion—yes, "bloke" is used 27 times more often in the British National Corpus (1980-93) than in the Corpus of Contemporary American English (1990-2015)—he expands the search to the "Harry Potter" series, then "Harry Potter" fan fiction, then online erotica, until he runs out of steam. His readers will have skipped ahead, maybe to the part about how an author's novels become longer as her books become more popular. It's true. Mr. Blatt can prove it.

"The adverb is not your friend," Stephen King wrote in "On Writing." Mr. Blatt attempts to prove this by running some tests on Hemingway's adverb rate. It's low: 80 adverbs ending in –ly (sleepily, irritably, sadly, etc.) per 10,000 words across 10 novels. Mr. Blatt follows up with charts that show that the most acclaimed novels by Hemingway ("The Sun Also Rises" and "A Farewell to Arms"), William Faulkner ("The Sound and the Fury" and "As I Lay Dying") and John Updike (the Rabbit Angstrom tetralogy) are those with the lowest adverb rates.

Mr. Blatt drills down to what could be considered an unreasonable level: He downloads more than 9,000 novel-length fan-fiction stories and concludes that amateur writers use –ly adverbs more than the pros do. The amateurs are wildly, gracelessly and spectacularly unsuccessful in their excess.

Guess what else? Men don't write about women the way that women write about men. Mr. Blatt finds that J.R.R. Tolkien uses the word "he" just under 1,900 times in "The Hobbit" and the word "she" only once. Elmore Leonard used "he" more than "she" in every one of his 45 novels. Joseph Conrad, Cormac McCarthy and Herman Melville did the same: more "he" than "she" in each of their novels. Surprisingly, the same is true for female novelists. Jane Austen is the only one among the unspecified number Mr. Blatt surveyed who used "she" more than "he" in every one of her books.

In 2001, Elmore Leonard famously advanced a list of 10 rules for writing. It included a prohibition on the word “suddenly” and a restriction of two or three exclamation points per 100,000 words of prose. Mr. Blatt tries to catch him breaking his own rules and finds that Leonard used “suddenly” often in his early westerns but not once in the nine novels he wrote after 2001. Same story with exclamation points: Leonard didn’t eliminate them, but his frequency was nowhere near that of Tom Wolfe (no surprise!) or Tolkien (who was also a “super-user” of “suddenly”).

It’s easy for Mr. Blatt to look for word patterns and unmask Mr. King or J.K. Rowling when they publish under pseudonyms. That rumor, first advanced by a radio talk-show host in 1976, about J.D. Salinger and Thomas Pynchon being the same person? No way. Word choice is constant, Mr. Blatt writes, and “authors do end up writing in a way that is both unique and consistent.” For one thing, Mr. Pynchon rarely uses the same word to begin two consecutive sentences. In his novel “Bleeding Edge,” only 1.6% of the more than 10,000 sentences begins with the same word as the previous sentence.

The results of a random set of Mr. Blatt’s experiments are enlightening—who would have thought that Nabokov’s favorite word was “mauve”?—but not always in the way he imagines. Algorithms are fast and accurate, but it turns out that literature can hold its own against those who try to break it into pie charts and probability graphs. Finding the magic takes more than a download. You have to read the book.

Mr. Baker is a writer and editor in Portland, Ore.

Appeared in the Mar. 15, 2017, print edition as ‘Algorithms Meet Adverbs.’