

Reading Between the Lines: Why Linguistic Analysis Matters in Literary Biography

When we read scholarly biography, we tend to trust it. Citations reassure us. Academic credentials reassure us. The authoritative tone—calm, measured, properly footnoted—reassures us most of all. We assume that if a scholar says someone was “vague and untrustworthy,” they must have evidence for it. We assume that when mysteries are declared (“The Oxford sojourn is shrouded in mystery”), the mystery exists in the archive, not in the telling.

But what if the unreliability isn’t in the subject—but in how the biographer writes about them?

This is where linguistic analysis becomes essential. Not as an attack on individual scholars, but as a method for revealing what sits beneath the surface of authoritative-sounding prose: the patterns, the framing devices, the rhetorical strategies that shape how we read a life.

The Method: Making the Invisible Visible

The process is straightforward. Take a chapter of scholarly biography—in this case, Islay Murray Donaldson’s Chapter 2 on S.R. Crockett’s Edinburgh years. Convert it to plain text format. Then analyze it systematically using established linguistic methods:

Attribution analysis: How does the author introduce the subject’s own statements? With neutral verbs (“states,” “reports”) or loaded ones (“boasts,” “exaggerates”)?

Modal verb analysis: When does the author express certainty versus uncertainty? Are they certain when making negative character claims but uncertain about actual facts?

Rhetorical question analysis: Where does the author use questions to plant speculation without requiring evidence?

Lexical choice analysis: What words are repeatedly associated with the subject? What connotations do they carry?

Passive voice analysis: When does the author remove their own agency, presenting interpretation as objective fact?

Evidence-to-assertion ratio: Which claims are supported by citations, and which are simply stated as true?

The beauty of this approach is that it’s quantifiable. You can count instances. You can create tables. You can demonstrate patterns that might feel instinctively wrong to a careful reader but are hard to articulate without systematic analysis.

This process can be done 'hand in hand' with AI analysis tools. It's useful (of course) for the human to have knowledge of linguistic analysis in order to direct and 'quality control' the output.

What the Numbers Reveal

In Donaldson's Chapter 2 (240 'txt' lines), the linguistic analysis reveals:

- **6 direct negative characterizations** of Crockett's reliability ("always exaggerates," "vague and untrustworthy")
- **0 citations** supporting those characterizations
- **5 attribution verbs** with negative connotations when introducing Crockett's statements ("boasts," "plumes himself," "show off")
- **3 rhetorical questions** suggesting dishonesty without requiring proof

By contrast, when Donaldson makes factual claims—dates, locations, institutional records—she provides proper scholarly citations. The reader unconsciously conflates the two types of statement, assuming both are equally evidenced. But they're not.

This isn't about conscious deception. Scholars are human. We develop rhetorical strategies, often unwittingly, to make our arguments more persuasive. We choose words that support our interpretations. We frame evidence in ways that guide readers toward our conclusions. Linguistic analysis doesn't judge these choices as morally wrong—it simply makes them visible.

Why This Matters Now

We live in an age when AI systems train on secondary sources and reproduce their patterns wholesale. When a large language model encounters Donaldson's statement that "Crockett always exaggerates," it doesn't distinguish between evidenced fact and rhetorical assertion. It absorbs the pattern: Crockett = unreliable. The next researcher who asks an AI about Crockett receives that characterization as established truth.

This is the 21st-century problem of inherited narrative, amplified by technology. Errors don't just accumulate—they multiply exponentially. A single unsupported assertion in a 1989 monograph can become the foundation for a thousand AI-generated summaries, each one repeating the pattern with increasing confidence.

Linguistic analysis offers a corrective. It asks: *Where's the evidence behind this claim? What rhetorical work is this language doing? How does the framing shape what we think we know?*

The Crockett Case: A Worked Example

Our linguistic analysis of Donaldson's Oxford section demonstrates the method's value. She writes: "It is not clear why he went to Oxford, or for how long he stayed, or what he did there... The Oxford sojourn is shrouded in mystery."

Pattern analysis reveals: - Lists of unknowns presented as suspicious rather than as normal archival gaps - Hyperbolic language ("completely unexplained and incomprehensible") -

Baseless speculation (could he have visited Margaret Oliphant?) immediately dismissed but left lingering - Passive construction (“is shrouded”) implying concealment rather than absence - Zero engagement with contemporary sources (Cassell’s Family Magazine, April 1895) that answer every question

The “mystery” exists in the rhetoric, not the archive. And the Cassell’s interview—hiding in plain sight for decades—proves it.

This matters because Donaldson’s characterization has been cited as authoritative. Later scholars inherited her framing. AI systems may now reproduce it. Consequently the pattern can spread, detached from evidence, creating a new and less reliable narrative.

Keeping Criticism Honest

Linguistic analysis isn’t about attacking individual scholars. It’s about holding the field to its own standards. If we claim to work from evidence, our language should reflect that. If we’re speculating, we should signal it clearly. If we’re making character judgments, we should provide evidence proportional to the claim.

The method helps us distinguish: - **What we know** (cited, documented, verifiable) - **What we interpret** (reasonable inference from evidence) - **What we assert** (rhetorical framing without evidential support)

In literary criticism, these boundaries often blur. We read patterns into texts. We develop arguments through close reading. We make interpretive leaps. None of this is inherently wrong—it’s how literary analysis works.

But biography operates under different rules. When we write about real people’s lives, evidence matters. When we characterize someone as dishonest, we need proof. When we declare mysteries, we need to acknowledge whether the mystery is archival or rhetorical.

Linguistic analysis helps us see when we’ve crossed those lines—or when previous scholars have.

The Broader Lesson

This method has applications far beyond the Crockett case. Any time inherited narratives feel suspect, any time authoritative prose makes claims without citations, any time character judgments appear detached from evidence—linguistic analysis can reveal the patterns beneath.

It’s particularly valuable now, when AI systems amplify whatever patterns exist in our scholarship. If we want future researchers—human or machine—to distinguish evidence from assertion, we need to make those distinctions visible in our own work.

The tables, the counts, the systematic analysis: these aren’t pedantry. They’re rigor made transparent. They’re a way of showing our work, of demonstrating not just *what* we conclude but *how* we got there.

And sometimes, they reveal that what looked like scholarly consensus was actually rhetorical construction—persuasive, influential, widely cited, but ultimately unsupported by the evidence it claimed to interpret.

That's not an attack on scholarship. It's an invitation to do scholarship better.

[Read the full linguistic analysis with 12 detailed tables.](#)

[Read a comparative linguistic analysis of Donaldson's Chapter 2 and Phillips' Oxford Problem Resolved](#)
