



A Much Bigger Bibliography of Johnsonian Studies

JACK LYNCH[†]

My latest project began as an experiment to find out whether the current generation of so-called artificial intelligence – the generative AI provided by large language models – could do anything useful for a fairly traditional scholar of eighteenth-century studies. It can be found on [my website](#).

Though I once had a hand in what was then known as humanities computing, and is now known as digital humanities, I haven't kept up. I know today's digital tools can do impressive things, but most of those things don't appeal to me. I'm not interested in coding, distant reading, network analysis, or visualization tools. I also have no desire, for reasons both practical and principled, to have AI write anything for me that I intend to present as my own. Nor do I want AI to do the hard work of reading for me. Engagement with texts is, at least in the kind of scholarship I want to do, irreplaceable, and I don't want proxies for texts. But if AI tools can bring me closer to texts I should read, particularly by sorting through the vast amount that's been published, then they may have real value.

This curiosity prompted me to return to a project I began in the previous millennium, an enumerative bibliography of scholarship related to Johnson that continued the series of bibliographies by James L. Clifford, Donald J. Greene, and John A. Vance between 1951 and 1987. The latest of those bibliographies covered up to the year 1985, so I began collecting citations from 1986 and later. My first contribution took the form of a long article in the pages of *The Age of Johnson* 10 (1999); the following year it appeared as a standalone *Bibliography of Johnsonian Studies, 1986–1998*, from AMS Press. Not long after that I made the whole thing available on the World Wide Web, and I kept it up to date through the mid-2010s.

[†] Jack Lynch is Distinguished Professor of English at Rutgers University, Newark. He is the author or editor of many books on Samuel Johnson and co-editor (with John Scanlan) of *The Age of Johnson*. He can be reached at [jlynch\[at\]newark.rutgers.edu](mailto:jlynch[at]newark.rutgers.edu).

But life got busy as I was dragged into university administration. The regular updates slowed, then eventually stopped altogether. Last year, though, I decided it was time to bring the old project up to date. I turned to all the library databases I had access to and downloaded citations to the last decade's worth of scholarly books and articles on Johnson — more than 1,500 of them.

The first question was how to integrate these new items into my larger bibliography. I relied on the citation manager Zotero, which I've been using to keep track of scholarly sources for some time. It's freeware, available for Windows, Mac, Linux, and mobile devices; it's easy to search and it can produce bibliographies in dozens of formats. Getting the new material into Zotero was easy: library databases spit out files in .bib (BibTeX) format, which is easily read by citation managers, so I quickly built a Zotero database containing the new records. But the old bibliography had been assembled by hand, as it were, and it resisted easy translation into consistent records. Free-form information, not divided into clearly demarcated fields, is a nightmare for databases, which struggle to distinguish chapters from articles, book series from journals, and volume numbers from issue numbers.

That was the first challenge I set for AI. I fed Google's Gemini a portion of my online bibliography and instructed it to convert it into .bib files I could import into Zotero. It obliged. It took a few tries to get everything right, but in the end it did an impressive job, saving me dozens of hours of tedious data entry. The 3,000 items in my old bibliography joined the 1,500 from new sources, I weeded out the duplicates, and it was up to date.

But why stop with my own bibliography? I got my hands on PDFs of the Clifford, Greene, and Vance bibliographies and asked Gemini to give them the same treatment as mine. The 5,500 items in those bibliographies (without any of the compilers' commentary) became .bib records and joined the 4,500 records in my database. And why stop with Johnson? I had always limited my coverage to items of specifically Johnsonian interest, including items related to James Boswell and Hester Lynch Piozzi only when they also touched on Johnson. This time, though, I expanded my scope: everything I was able to find related to Johnson, Boswell, or Piozzi made it in. The Boswellian bibliographies of Brown and Cochrane got the same treatment as Clifford, Greene, and Vance. There is no comparable bibliography of Piozzi, but I collected everything I could and integrated

that as well. The result was an omnibus bibliography of more than 10,000 items of Johnsonian, Boswellian, and Piozzian publications from 1738 to 2026.

The next step was seeking out older items that had escaped the attention of earlier bibliographers, including myself. A great deal of old material has been cataloged in recent decades, including countless eighteenth- and nineteenth-century newspapers and magazines. I pulled it all together, converted it to .bib files, and watched as the Zotero database grew.

The result is a bibliography of more than 27,500 items, including around 2,300 books and pamphlets, 2,700 book chapters, 5,800 newspaper articles, 1,200 magazine articles, and 14,700 journal articles, of which about 6,700 are book reviews. About 1,700 come from the eighteenth century, 4,100 from the nineteenth, 16,400 from the twentieth, and 5,200 from the first quarter of the twenty-first. They range in length from a three-line letter to the editor to the three million words of Johnson's *Dictionary*.

A bibliography of this size has obvious advantages, but the needle-in-a-haystack problem is real: how are we supposed to know what's *in* those 27,500 items? Some are helpfully titled — we have a pretty good idea of what kinds of information we'll find in an article titled "Was Dr. Johnson a Smoker?" or "A Copy of Shakespeare's Works Which Formerly Belonged to Dr. Johnson." But what to do with the many hundreds with titles like "Boswell Up-to-Date," "Dr. Johnson Discomfited," "The Greatest Thinkers," or "Anecdote of Dr. Johnson"? What to do, heaven help us, with the 250 items titled simply "Dr. Johnson," or the 450 titled "Samuel Johnson"? They've been sitting in libraries, indexed but presumably unread, for generations. How many contain things we might care to consult?

Clifford, Greene, and Vance addressed this problem with a numerical taxonomy of their own devising: heading 1 covered "Bibliography"; heading 2 was "Editions"; heading 3 was "Biography," with "Boswell" getting his own heading 4; and so on. Some top-level headings have subheadings, so under category 1 ("Bibliography") we find 1/1 ("Johnson's Writings"), 1/2 ("Writings about Johnson"), 1/3 ("Books Owned by Johnson"); under 10 ("General Comment") came 10/1 ("Johnson as Critic"), 10/2 ("Johnson as Biographer"), 10/3 ("Johnson as Conversationalist"), and so on. In each of these categories, individual items — books, chapters, articles, reviews — were numbered sequentially by date of publication: under 11/8 ("Women and Marriage") came 11/8:1, W. H. Craig's *Dr. Johnson and the Fair Sex*

(1895), then 11/8:2, G. Valbert's "Le docteur Samuel Johnson et les femmes" (1896), and so on, right up to 11/8:20, R. P. Rama's "Johnson and Rousseau on the Woman Question" (1985).

It's not a bad system. If I'm seeking commentary on *Taxation No Tyranny*, I can browse category 20, "Political and Economic Writings and Views"; if I come across a reference to William B. Todd's "Concealed Editions of Samuel Johnson" (1953) and want to know what it's about, I can look up Todd's name in the Clifford-Greene index and discover it belongs to category 20, and is therefore about the political writings. But the system was, inevitably, limited. Clifford and Greene, for instance, classify Louis Kronenberger's "Johnson and Boswell" and Charles Whibley's "Samuel Johnson: Man of Letters" under the heading 10/6, "General Assessments of Johnson and Miscellaneous Comment," but that's hardly informative; it tells us only that they're among the 396 items too miscellaneous to appear anywhere else. The categories also showed their age: it felt positively condescending to put serious modern scholarship from the 1990s under "Johnson's Views: Women and Marriage." And as scholarship in the 1980s and '90s moved away from studies of single works and single authors, it became increasingly difficult to slot books and articles into any categories at all.

In my late-'90s contribution to the series, therefore, I reluctantly decided to forgo the taxonomy, convinced there was a limit to how much useful work the catchall category 10/6 could do. I decided that the basic search functions provided by web browsers would do at least some of the work of classification: a search for "Rambler" or "Thraliana" or "Jacobit_" would let users find things likely to be of interest. Whenever possible, if authors provided abstracts or publishers provided blurbs, I transcribed them. And sometimes I added very brief abstracts or descriptions of my own composition to the online bibliography. I curtly described Freya Johnston's superb *Samuel Johnson and the Art of Sinking, 1709–1791* (2005) as "A learned meditation on Johnson's interest in 'littleness.'" To John Geirland, "Doctor Feelgood: Stricken by 'Vile Melancholy,' the 18th-Century Critic and Raconteur Samuel Johnson Pioneered a Modern Therapy," published in *Smithsonian* magazine in 2007, I appended, "A brief biographical overview, with an argument that SJ's attempts to ward off his melancholy anticipated modern cognitive-behavioral therapy." It wasn't much, but it was something.

But with 27,500 items, including large numbers from unscholarly sources like newspapers and magazines, the problem becomes much more acute. Not nearly enough of the titles were usefully descriptive, only a tiny number have authorial abstracts, and there was simply no way I could hope to read and summarize 27,500 items. (Have I read 27,500 scholarly books and articles in my life?) And so I again turned to Google's Gemini. I fed it PDFs of publications and asked for summaries. At first I considered using it to extract just some searchable keywords, but I was unhappy with the results. In the end I decided that abstracts would be more useful.

It took a lot of effort to get useful abstracts out of Gemini. My concern was always to produce an easily searchable description that could tell a scholar enough about a source to decide whether it's worth consulting it. Eventually I settled on a set of prompts that yielded abstracts I liked, and I applied it to all the electronic texts I could get my hands on. At times it felt a bit like a digital assembly line — downloading books and articles by the hundred, feeding them into Gemini for abstracting, pasting the generated abstracts into Zotero — but I was able to do it in the spare time of a few months. The sheer volume made it time-consuming, but each individual item went quickly: it might take just twenty or thirty seconds to abstract a scholarly article and copy the abstract into Zotero. The result is a little more than 24,000 abstracts — that is, around 87 percent of the 27,500 items in the bibliography come with automatically generated abstracts. (The remaining 13 percent are not in any electronic form I had access to.)

I should emphasize that these generated abstracts are really abstracts, not evaluations. They've been done almost entirely without my intervention. They're machine-generated, and because they merely reprocess the authors' words, they can offer no opinions on whether the argument's sound. An article can be entirely wrongheaded, poorly written, and backed up by terrible evidence; it will not be called out, just dutifully summarized. That's probably just as well, because I don't think we should trust AI for judgments quite yet.

I recognize that many are not merely skeptical of AI tools but openly hostile to them on multiple grounds, and the industry's contempt for intellectual property is high on many people's lists. I therefore set the software not to use any material I provided for its own training purposes, and not to retain anything I had uploaded — it's all treated as confidential. Though I suspect the software has already been trained on everything I have access to, I've not fed the AI beast on modern Johnsonian scholarship.

Those skeptical about AI are always quick to ask: are the abstracts any good, and how can I know that? I freely admit I haven't read more than a tiny fraction of them. But I tested the system on my own writing, and on works I know intimately, and found Gemini generally did a pretty good job of summarizing. Sometimes it misses nuances, and it's immune to irony. But most of the abstracts are good enough.

So-called AI "hallucinations" are, of course, a real danger. AI, notoriously eager to please, churns out nonsense when it doesn't have good answers. But I discovered that, while hallucination is all too common when you ask it about the world at large, it's rare when you provide it with text and ask for a summary. Only a few times did I catch Gemini making outlandish mistakes, and they were almost always because I had inadvertently failed to provide it with the necessary text. I came across an article from 1916, for instance, called "Where Johnson Made a Blunder." It's about his disapproval of Robert Mayhew's designs for Westminster Bridge, and the *Westminster Gazette* thought, in retrospect, it was an obvious blunder. But, because the OCR in the library database had gone wrong, the text I fed to Gemini contained the title but was missing the text of the article. When asked to summarize an article called "Where Johnson Made a Blunder" with no additional information, Gemini dutifully obliged with an abstract of 125 words on famous mistakes in the *Dictionary* (*pastern*, *windward* and *leeward*, and so on). It was a pure digital fantasia. I happened to catch that one, but I'm sure others have slipped past me. Still, I consider them a small price to pay for reasonably accurate summaries of far more books and articles than any human could hope to read. The point, after all, is not to take the place of the original publications, but to point scholars toward them.

The resulting bibliography, with its 27,500 items and 24,000 abstracts, is vast. The full text is around 3.6 million words, which makes it about as long as the *Dictionary*, the *Rambler*, and the *Idler* put together, pretty much everything Johnson published in the 1750s. That makes it far too long ever to appear in print. I've put the entire thing up as a single web page, but web browsers strain under its weight. Much better than using the web page is to download the databases from which it was generated — they're linked at the top of the web page — and to use Zotero or another citation management system. The databases include a few things not visible on the web page, including the reference numbers from the bibliographies by Clifford, Greene, Vance, and Cochrane, and sometimes provide additional

information about later publications. Zotero makes it easy to pull up every item that contains a reference to the *Harleian Miscellany*, every newspaper article that mentions the Inner Temple, or every biographical item that touches on Wales.

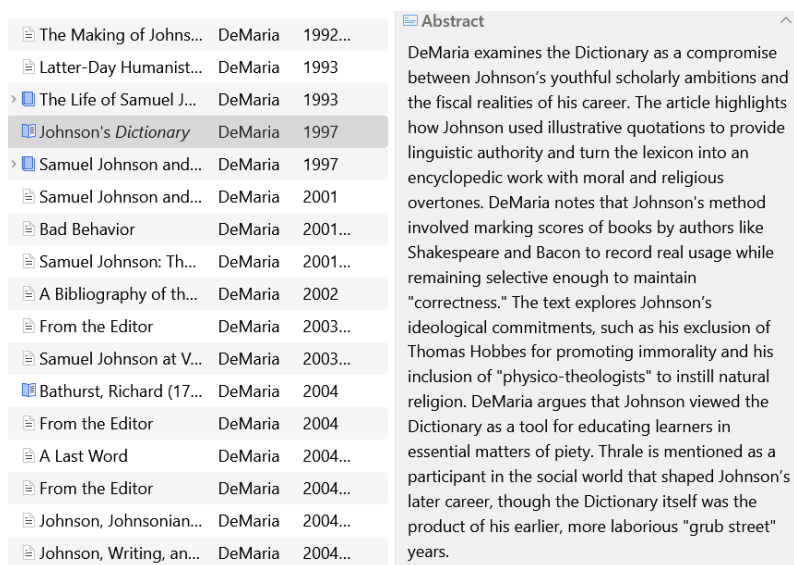


Figure 1: This screenshot of Zotero shows the AI-generated abstract for “Johnson’s Dictionary,” a 1997 article by Robert DeMaria, Jr.

There’s one more experimental element to the project. Out of curiosity I took the entire bibliography, all 3.6 million words of it, and fed it back into another AI tool by Google, NotebookLM (LM for *language model*). This works like a typical AI chatbot, but with a crucial difference: it draws its answers only from a library of texts the user provides. It becomes, in other words, an expert on the subject of its database. This makes it possible for anyone to ask plain-language questions about the bibliography itself, things that go beyond what can be accomplished even in a fairly complex database search. Zotero can instantly tell me which items mention *Troilus and Cressida*, but is at a loss with questions like “In which twentieth-century dramas does Johnson appear as a character?”

NotebookLM, though, is excellent at this. Ask it a question about a topic addressed in the bibliography and the abstracts, and it produces a pretty good answer. “Describe the trajectory of S.C. Roberts’s interest in Johnson over the course of his career.” “How do Scottish and English newspapers differ in their response to the discovery of the Boswell papers?” “What effect did second-wave feminism have on the study of

Piozzi?” Its greatest strength is generalizing: ask it, “What were the major trends in Johnsonian scholarship in the early 1980s?” and it’ll tell you about the conferences and publications for the bicentenary, the questioning of Boswellian authority, the focus on Johnson’s legal thought and his devotion to social reform and humanitarianism, the introduction of post-structural theory to eighteenth-century studies, and so on. Ask it, “Give examples of people invoking Johnson’s authority to address political questions in their own day,” and it starts with Wesley’s republishing *Taxation No Tyranny* during the American Revolution, continues through Thomas Gisborne and William Wilberforce’s abolitionist invocations of the argument for Joseph Knight, the use of Johnson as a champion of Irish independence by followers of Charles Stewart Parnell, the rival citations of Johnson on women’s rights and capabilities by suffragists and anti-suffragists, a Great Depression–era quotation from the *Journey to the Western Islands* about how the “rage of trade” would collapse once all nations became traders, the warning in *Rasselas* that teaching men to fly would allow “northern savages” to drop destruction from “an army sailing through the clouds” that became disturbingly relevant during the Blitz... The answers are all drawn from the bibliography, and every assertion, nearly every sentence, is backed up with one or more links to the relevant scholarship.

The bibliography is, inevitably, a work in progress, and I expect to continue tinkering. I hope people find it useful, and I welcome additions, corrections, and reactions of all kinds.