

(<https://blog.blakearchive.org/>)

The Blog of the Blake Archive and Blake Quarterly

BLAKE QUARTERLY, DIGITAL HUMANITIES

“Swift winged words”: the vocabulary and word distribution of Blake/*An Illustrated Quarterly*

APRIL 1, 2016

In the process of preparing the team to correct spelling errors in the digital archive of *Blake/An Illustrated Quarterly* (errors in the transcription and in the original print version), I have made a few interesting observations about word distribution in the *B/Q* corpus (before the first online issues). On the principle that misspellings occur less frequently than correct spellings, I used a series of PHP scripts to generate a wordlist sorted by the number of instances of each unique word. The team is beginning by checking the spellings of rare words (appearing 1-3 times in the whole print-only run of *B/Q*), in hopes of encountering a higher percentage of errors more efficiently. The word list itself, however, tells us some interesting things about the word choices of Blake scholars.

A few caveats about the following figures: First, this list was constructed to check spelling, so different spellings (British and American, not to mention the idiosyncratic spelling of Blake and other authors) are counted as separate words, and—because capitalization is sometimes important to correct spellings as well—different capitalizations of the same word are also counted as separate words. Second, metadata (e.g., the journal title, article titles, editors, and contributing authors) has been

included in this word count as well as transcribed text—so the word “Quarterly,” for instance, is counted more than it would be in the transcription alone. Third, although clusters of numbers and symbols (such as “\$1,000”) are not counted as “words,” Roman numerals and combinations of numbers and letters (such as “65r”—i.e., the recto side of page 65) are counted. Fourth, most words were searched using a regular expressions match that recognized word breaks, but some had to be searched using a less accurate tool that accommodated special characters. Fifth, these figures shift slightly as we correct spellings.

On the one hand, *BIQ*'s word usage is in some ways like that of ordinary English. Its most common words—the thirty-four words which appear more than 10,000 times each—are either simple Anglo-Saxon words of a single syllable or part of Blake's name: the, of, and, in, to, a, Blake, is, The, that, by, as, for, on, with, Blake's, his, from, not, it, are, was, William, be, at, which, A, this, or, an, I, have, he, but. These most common words constitute nearly a third of all the word instances (32.74%), the word “the” alone accounting for every twentieth word (5.29%), while “William,” “Blake,” and “Blake's” together account for two out of every hundred (2.18%).

On the other hand, *BIQ* employs a very large vocabulary: 110,452 unique “words” when distinguishing different capitalizations and spellings. If we exclude words containing diacritics, non-Roman alphabetic characters, unusual punctuation, and Arabic numerals—that is, if we count only words using the core Roman alphabet (A-Z without diacritics), periods, and apostrophes—the number of unique words drops to 97,496. If we take that set and ignore differences in capitalization, the number shrinks to 76,212. This still includes some non-English words, Roman numerals, and many proper nouns, and includes variations of spelling, but it also excludes items which are arguably English words: “encyclopædia,” “naïveté,” “1800s,” “3rd,” “&c.” Even the very conservative count of 76,212 indicates a gigantic vocabulary, many times the famously-large vocabulary of Shakespeare.

Estimates of the size of Shakespeare's vocabulary vary from about 15,000 words (F. Max Müller) to 29,066 (Marvin Spevack) in part because the latter counted “every different word inflection found, not just each word's root.” We are counting even more expansively because we are distinguishing different spellings as different “words,” and are not restricted (like Shakespeare, for the most part) only to English words. Even excluding the foreign alphabetic characters of Greek and Hebrew, *BIQ*'s words written with the core Roman alphabet include words in (not an exhaustive list) Latin, French, German, Italian, Spanish, Catalan, Japanese (transliterated), Danish, Norwegian, Dutch, Polish, Russian (transliterated), Gaelic, and Middle English. Some unique “words” are really spelling variants, or referents (like “A” for a particular copy of one of Blake's works). But a considerable number of them are distinct words.

Of the 4.7 million word instances—and for these figures we are again including “words” using non-Roman characters, numbers, and symbols—usage distribution is broken down as follows:

Words each appearing 10,000 – 250,130 times (34 unique words)	1,547,184 instances (32.74% of total)
Words each appearing 1,000 – 9,999 times (551 unique words)	1,313,351 instances (27.79% of total)
Words each appearing 100 – 999 times (4,000 unique words)	1,119,637 instances (23.69% of total)
Words each appearing 10 – 99 times (18,045 unique words)	543,047 instances (11.49% of total)
Words each appearing 1 – 9 times (87,822 unique words)	202,303 instances (4.28% of total)

Some of the most interesting words are at the rarest end of the spectrum: *BIQ*'s hapax legomena, words which each appeared only once between 1967 and the last print-only issue in 2011. Over 46 thousand words appear only one time. Again, these “words” would not all appear as separate entries in a dictionary, or appear in a dictionary at all. 3,643 of them begin with numbers (“0.12s.6d,” “1/4th,” “120r,” “244mm,” “43ff,” “8ober,” “92nd,” etc.), not to mention those that begin with a letter and contain numbers, like “B875.” Some of the words are not English, and a few are not even in standard Roman letters—125 begin with extended Roman characters like å, Œ, and ž, 110 are Greek, and 52 are Hebrew. Some words are capitalized unusually: “BRAIN” appears only once in all capitals, but 30 times as “Brain” and 79 as “brain.” Yet others have bracketed portions: “B[r]other,” “B[rothe]r,” and “B[rother]” all appear once each (counted separately from “Brother” and “brother,” not to mention variants of “brothers” and “brother’s”).

If we count only “words” spelled with periods, apostrophes, and letters A-Z without diacritics, and ignore capitalization, there are still 27,679 hapax legomena. Some are merely unusual spellings, like “perswade” (quoted in Robert N. Essick’s review of G. E. Bentley, Jr.’s *Blake Books*, *BIQ* 11.3) or (punningly altering the word *neglect* to match the name of Gleckner) “neGLECK” (in a poem by Donald H. Reiman, *BIQ* 9.2). Many, however, are English words that really appear only once. Some of these words would be commonplace in other contexts, and are rare only because this is academic prose: “picnic” appears only in a quotation from Emerson (Morton D. Paley, “A New Heaven is Begun,” *BIQ* 13.2). Some words reflect critical approaches, like the Derridean term “diaphoristics” in Nelson Hilton’s review of David Punter’s *Blake, Hegel and Dialectic* (*BIQ* 17.4). Other words reflect attention to the material concerns of a craftsman like Blake—like the mention (in Joseph Viscomi’s “Blake in the Marketplace 1852,” *BIQ* 29.2) of paper manufacturers pressing or “calendering” paper. Some technical terms are used metaphorically: Swedenborg uses “cohobation” (redistillation) as a metaphor for spiritual purification (quoted in Peter Otto’s “A Pompous High Priest,” *BIQ* 35.1), and Frank Graziano, in his poem “The Last Judgement” (*BIQ* 12.1) applies the medical term “disphasia” (a neurological speech disorder) to a book. Other words are more homely in meaning but no less uncommon: a “bothy” (cottage) appears in Keri Davies’s “William Muir and the Blake Press at Edmonton” (*BIQ* 27.1),

while “drek” (a variant spelling of dreck, the latter spelling never occurring at all in *BIQ*) appears in Morris Eaves’s review of Abby Robinson’s novel *The Dick and Jane* (*BIQ* 21.1), in a quotation from the novel.

Eaves’s review of *The Dick and Jane* also provides other *BIQ* hapax legomena: “Blakeanity,” “Blakette,” and “Blakaliens.” Other Blake-based terms that appear only once include nouns such as “Blakeophile” (David Worrall, review of Joan Evans, *A History of The Society of Antiquaries*, *BIQ* 11.4), “Blakers” (quoted from Diana Hume George’s *Blake and Freud*, in a review by Thomas A. Vogler, *BIQ* 16.2), “Blaker Babe” and “Blake-o-holic” (t-shirts quoted in Robert N. Essick’s “Blake in the Marketplace, 2009,” *BIQ* 43.4), adjectives including “Blakeocentric” (Mary Lynn Johnson, “Blake’s Engravings for Lavater’s *Physiognomy*,” *BIQ* 38.2), “Blakesque” (William Muir, quoted in G. E. Bentley, Jr.’s “‘Blake . . . Had No Quaritch,’” *BIQ* 27.1), and “Blakeless” (Joseph Viscomi, “A ‘Green House’ for Butts?” *BIQ* 30.1), and even verbs such as “Blakespotting” and “Blaking” (in the titles of articles by Mike Goode and James Chandler, cited in G. E. Bentley, Jr.’s “William Blake and His Circle,” *BIQ* 41.1). *BIQ*’s vocabulary thus includes both the lexical richness of academia and the lexical richness of creative works—not only the works of Blake but the works of modern-day poets, novelists, and t-shirt designers.

by Adam McCune

PREVIOUS

Virtual Blake

NEXT

Revenue Stamps on Receipts