Computational Humanities Research

www.cambridge.org/chr

The quotable canon: Poetry as example in the Princeton Prosody Archive (1532–1929)

Wouter Haverals[1], Meredith Martin[1,2], Rebecca Sutton Koeser[1] and Laure Thompson[1]

[1]Center for Digital Humanities, Princeton University, USA and [2]Department of English, Princeton University, USA

Research Article

Cite this article: Haverals Wouter, Meredith Martin, Rebecca Sutton Koeser and Laure Thompson. 2026. "The quotable canon: Poetry as example in the Princeton Prosody Archive (1532–1929)." *Computational Humanities Research*, **2**:e28, https://doi.org/10.1017/chr.2026.10048

Received: 20 July 2026 UTC
Revised: 20 July 2026 UTC
Accepted: 10 September 2026 UTC

Keywords:
Bayesian hierarchical modeling; canon formation; historical prosody; poetic reception; text-reuse detection

Corresponding author: Wouter Haverals; Email: wouter.haverals@princeton.edu

Abstract

This article studies the afterlife of poems as examples. Using the Princeton Prosody Archive and a large reference corpus of anglophone poetry, we recover likely poem quotations across works on prosody, elocution, grammar, rhetoric, and literary instruction from 1532 to 1929. We treat reception as a practical history of availability: the ways poems were cut into lines, couplets, and stanzas and made to do explanatory work. Methodologically, the article combines text-reuse detection with corpus filtering and hierarchical modeling to distinguish spurious textual overlap from historically plausible quotation and to trace poem-level trajectories through the archive. Those trajectories reveal specific contours of a shift in the repertoire of teachable verse: the Augustan classical-translation apparatus that had organized verse pedagogy through the eighteenth century gave way, across the long nineteenth century, to a Romantic and lyric repertoire whose lines could travel without contextual framing. The resulting account describes the quotable canon as a collection of poems and fragments whose authority was made through use. It shows how poetic reception can be studied at scale while remaining attentive to the instability of quotation, attribution, and reuse.

Plain language summary

This article studies how poems were used as examples in books about poetry, pronunciation, grammar, and literary instruction. This means that we have to follow the smaller pieces of poems – lines, phrases, and stanzas – that teachers and writers quoted to explain how verse worked. We use the Princeton Prosody Archive, a large collection of historical books about poetry and language, together with a large database of English-language poems. Text-reuse methods help us find likely quotations, and statistical models help us trace how often poems appeared over time. The article shows that poetic authority was often made through use: poems became important because their lines could be copied, taught, remembered, and reused. Across the long nineteenth century, older classical and Augustan examples gradually gave way to a more Romantic and lyric repertoire. The result is a history of poetry's practical afterlife: how poems lived on in the ordinary books that taught readers what poetry was.

Introduction

A poem's afterlife often begins when its lines are called into service. In manuals, handbooks, and treatises on verse, scholars and teachers reach for borrowed passages so readers can hear a rhythm, follow a comparison, or settle a definition. Such borrowings do double work. They illustrate a local point about how poetry sounds and scans, and at the same time, they certify what actually counts as usable, authoritative wording for the point being made. Thomas Gray's *Elegy Written in a Country Churchyard* (1751) had barely settled into the republic of letters before it started doing this kind of secondary labor. Its opening stanzas appear in *Sheridan's and Henderson's practical method of reading and reciting English poetry* (1796) to do exactly what the title promises: drill students in the enunciation of long vowels and a slow, solemn delivery (Figure 1, top left). Half a century later, a mid-Victorian primer on the music of speech (1856) sets the *Elegy*'s opening stanza to staff notation to teach common time meter (Figure 1, top right). A literary history of the same period invokes it as a textbook case of plagiarism, tracing the opening line back to a passage in Dante's *Divine Comedy* (Figure 1, middle left). Stopford Brooke, in *Theology in the English Poets* (1880), pulls a different couplet and presses it into an argument about Pope and the eighteenth century's turn toward rural feeling. In the process, Brooke silently rewrites the lines: Gray's plural "they kept [...] their way" becomes a singular "has kept [...] his way" (Figure 1, middle right). By the end of the century, the poem has been broken up further still. W. W. Skeat's *Scansion of English Poetry* (1898) lists 32 metrical archetypes and illustrates them with single lines plucked from across the *Elegy* (Figure 1, bottom left).

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

The same poetry-as-scrapyard move appears in an early-twentieth-century English grammar, where a few words of a single line are lifted from their stanza – leaving the iambic pentameter mutilated – and recruited into an exercise on the use of adjectives (1906, Figure 1, bottom right).

Poetry is everywhere in the writing that teaches readers how to read verse, but rarely as itself. It enters these books as a resource, a store of lines familiar enough that an author can reach for one and trust it to make an instruction concrete. Once this happens, the poem begins to live at two scales. The title and author still lend authority, but the work of explanation falls to a smaller unit – a phrase, a line, and a couplet – that can be moved around. Guillory has called this the condition under which the canonical poem operates: its staying power depends less on qualities recoverable by close reading than on its success as a store of quotable phrases (Guillory 1993). Of Gray's *Elegy* in particular, Guillory argues that its popularity is largely due to the effect of turning "classical literacy into an anthology of quotable vernacular phrases." Figure 1 shows that process in its everyday mechanics for one well-traveled poem, but the logic is general: a poem's authority is sustained through the game of cutting and grafting (Compagnon 1979). A line is removed from one frame and stitched into another, where it serves a purpose its author did not intend.

This article studies some of those traces of use. We first explain how we recover likely poetry quotations in the Princeton Prosody Archive (PPA) and attach them to identifiable anglophone poems where the evidence allows. We then ask what that record of poetic traces can teach us about how poems move through time in this kind of writing. Some poems become anchors in prosody writing; others surface around a particular period; others again appear only briefly. By following these trajectories, we treat reception as a practical history of availability. We build a pipeline for recovering quoted poems at scale, assemble a dataset of citation events in the PPA, and ask what a poem's presence in the record of prosody writing can tell us about how poetic authority is made and sustained. Which poems entered this economy of borrowing? And what do their trajectories through it look like?

Finding poetic needles in the PPA haystack

The Princeton Prosody Archive

The PPA collects the scattered history of poetic instruction. The six Gray examples in Figure 1 come from this archive, which gathers the long, uneven record of English-language thinking about prosody, versification, and pronunciation. The PPA grew from the bibliographical and historical work behind Martin's *The Rise and Fall of Meter* (2012), itself shaped by T. V. F. Brogan's *English Versification, 1570–1980* (1981), and is now a full-text searchable collection of public-domain materials drawn from HathiTrust, Gale/ECCO, and EEBO-TCP (Martin et al. 2018). Its current dataset contains 7,122 works, ranging from Leonard Cox's *The art or crafte of rhetoryke* (1532) to I. A. Richards's *Practical Criticism* (1929). The center of gravity is the nineteenth century, when questions of meter, pronunciation, and literary instruction converged on the quoted line (Prins 2011). That makes the PPA especially useful for studying poems as examples: it preserves a print world in which verse was repeatedly broken into usable fragments and made to carry the work of explanation.

That usefulness does not make the evidence easy to collect. Even though the PPA has a powerful search *interface*, a quoted poem is not a stable search *object*. Quotation loosens the poem from the signs by which we would usually find it. Even for a poem as familiar as Gray's *Elegy*, searching for its title, first line, or author name surfaces only a fraction of its traces. In the PPA, poetry washes ashore in bits and pieces.[1]

Text-reuse detection

Digital humanities research has built up a tradition of computationally tracing how texts circulate and leave evidence in subsequent writing. The tools built for this task generally travel under the umbrella of "text-reuse detection" (though the label covers practices different enough that it can obscure as much as it reveals). Early computational work was highly targeted: researchers had to know roughly *where* to look and *what* to look for. For example, working on Virgil and Lucan, Coffee et al. (2012) searched for exact or near-exact word matches between specific, pre-selected classical texts. Their method systematically recovered the *loci similes* that commentators had long identified through close reading, now at scale. While powerful, this approach was also constrained: it required researchers to manually select the two texts for comparison, assume rough passage boundaries, and rely on relatively clean transcriptions for word-level matching to succeed. Yet, even within these constraints, Coffee et al. clearly envisioned the horizon of this methodology, noting that "digital approaches [...] can accelerate this research and provide new large-scale perspectives on intertextuality" (Coffee et al. 2012).

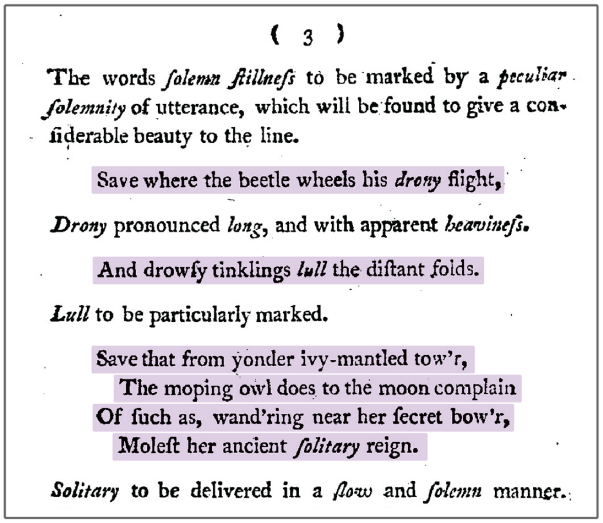
The shift they anticipated arrived with tools designed to scale up open-ended detection, such as *TextPAIR* (Gladstone, Horton, and Olsen 2008; Olsen, Horton, and Roe 2011), *TRACER* (Büchler 2013; Büchler et al. 2014), *textreuse* (Mullen 2015; Mullen and Li 2026), and *passim* (Smith et al. 2014). *Passim*, which we use in this article, was developed in the context of the *Viral Texts* project to discover reprinted passages across a large corpus of nineteenth-century American newspapers.[2] The tool was built for unstructured discovery: it is robust to OCR noise, relatively indifferent to passage boundaries, and requires no prior hypothesis about which texts will align. For *Viral Texts*, the method changed the scale of the question. Using *passim*, Cordell (2015) could follow reprinted passages across antebellum newspapers and show that the most widely circulating texts were often invisible to scholarly methods organized around named authors and titles. The newspaper exchange network had its own literary record, and text-reuse detection made that record available for study.

To reveal the similarly invisible labor of poetic quotation, we adapted this computational approach to a structurally asymmetric problem. Where *Viral Texts* tracked the horizontal circulation of articles within a single corpus of newspapers, recovering poems from inside prosody writing requires us to coordinate two distinct textual domains: the PPA, in which quotations appear, and a reference corpus of poetry against which they can be matched. We know, in broad terms, what the PPA contains. Shakespeare is quoted, and Milton, and – as Martin has shown (2019) – Elizabeth Barrett Browning. What we do not know are the specifics of those quotation histories: when particular histories of citation began, which obscure manuals first reached for them, or what poems beyond the poetic big guns turn up at all. The PPA contains nearly two million pages. To find these traces in a haystack of that

[1] For a related account of poetic authority in the PPA, see Haverals et al. (forthcoming), who trace how often poets are mentioned and which poets appear together.

[2] https://viraltexts.org/ (accessed June 3, 2026).

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.



(3)

The words *ſolemn ſtillneſs* to be marked by a *peculiar ſolemnity* of utterance, which will be found to give a conſiderable beauty to the line.

Save where the beetle wheels his *drony* flight,

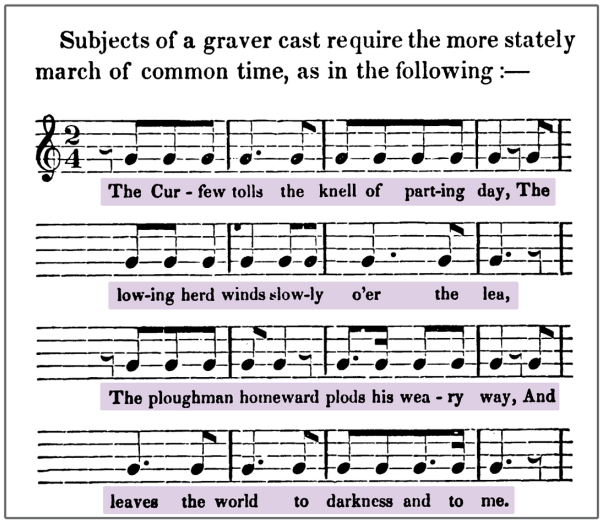
Drony pronounced *long*, and with apparent *heavineſs*.

And drowſy tinklings *lull* the diſtant folds.

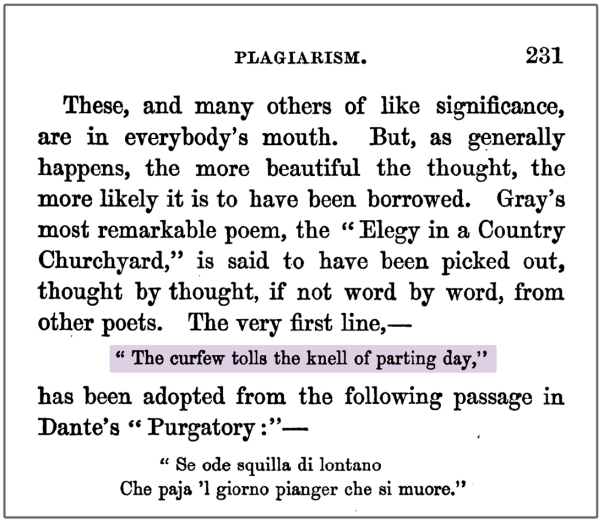
Lull to be particularly marked.

Save that from yonder ivy-mantled tow'r,
The moping owl does to the moon complain
Of ſuch as, wand'ring near her ſecret bow'r,
Moleſt her ancient *ſolitary* reign.

Solitary to be delivered in a *ſlow* and *ſolemn* manner.



Subjects of a graver cast require the more stately march of common time, as in the following :—



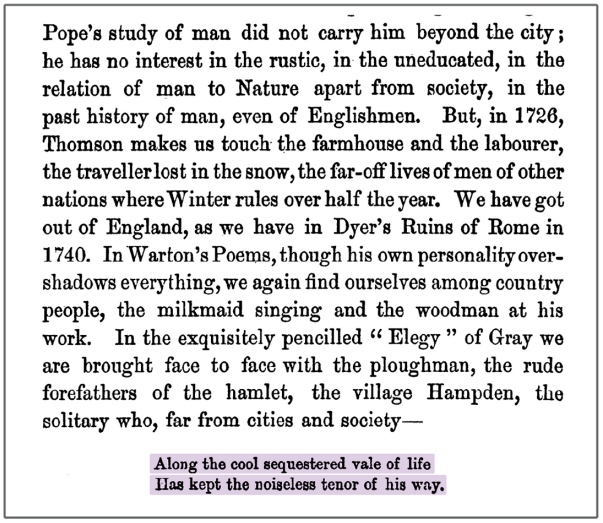
PLAGIARISM. 231

These, and many others of like significance, are in everybody's mouth. But, as generally happens, the more beautiful the thought, the more likely it is to have been borrowed. Gray's most remarkable poem, the "Elegy in a Country Churchyard," is said to have been picked out, thought by thought, if not word by word, from other poets. The very first line,—

"The curfew tolls the knell of parting day,"

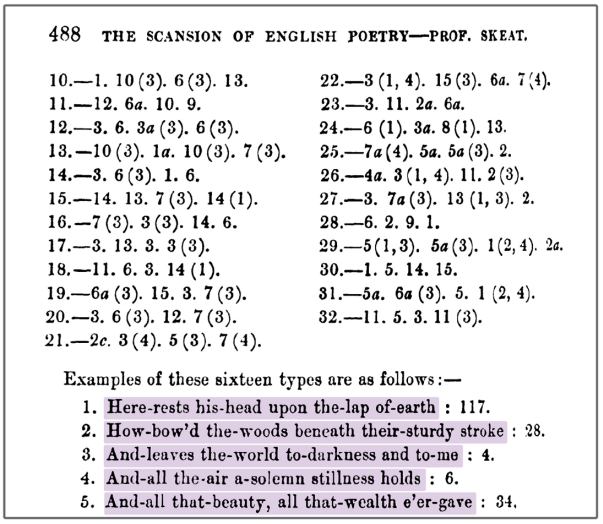
has been adopted from the following passage in Dante's "Purgatory :"—

"Se ode squilla di lontano
Che paja 'l giorno pianger che si muore."



Pope's study of man did not carry him beyond the city; he has no interest in the rustic, in the uneducated, in the relation of man to Nature apart from society, in the past history of man, even of Englishmen. But, in 1726, Thomson makes us touch the farmhouse and the labourer, the traveller lost in the snow, the far-off lives of men of other nations where Winter rules over half the year. We have got out of England, as we have in Dyer's Ruins of Rome in 1740. In Warton's Poems, though his own personality overshadows everything, we again find ourselves among country people, the milkmaid singing and the woodman at his work. In the exquisitely pencilled "Elegy" of Gray we are brought face to face with the ploughman, the rude forefathers of the hamlet, the village Hampden, the solitary who, far from cities and society—

Along the cool sequestered vale of life
Has kept the noiseless tenor of his way.

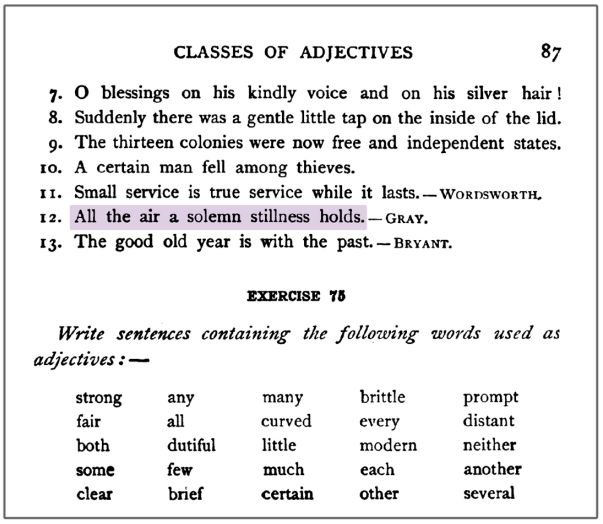


488 THE SCANSION OF ENGLISH POETRY—PROF. SKEAT.

10.—1. 10 (3). 6 (3). 13.
11.—12. 6*a*. 10. 9.
12.—3. 6. 3*a* (3). 6 (3).
13.—10 (3). 1*a*. 10 (3). 7 (3).
14.—3. 6 (3). 1. 6.
15.—14. 13. 7 (3). 14 (1).
16.—7 (3). 3 (3). 14. 6.
17.—3. 13. 3. 3 (3).
18.—11. 6. 3. 14 (1).
19.—6*a* (3). 15. 3. 7 (3).
20.—3. 6 (3). 12. 7 (3).
21.—2*c*. 3 (4). 5 (3). 7 (4).
22.—3 (1, 4). 15 (3). 6*a*. 7 (4).
23.—3. 11. 2*a*. 6*a*.
24.—6 (1). 3*a*. 8 (1). 13.
25.—7*a* (4). 5*a*. 5*a* (3). 2.
26.—4*a*. 3 (1, 4). 11. 2 (3).
27.—3. 7*a* (3). 13 (1, 3). 2.
28.—6. 2. 9. 1.
29.—5 (1, 3). 5*a* (3). 1 (2, 4). 2*a*.
30.—1. 5. 14. 15.
31.—5*a*. 6*a* (3). 5. 1 (2, 4).
32.—11. 5. 3. 11 (3).

Examples of these sixteen types are as follows :—

1. Here-rests his-head upon the-lap of-earth : 117.
2. How-bow'd the-woods beneath their-sturdy stroke : 28.
3. And-leaves the-world to-darkness and to-me : 4.
4. And-all the-air a-solemn stillness holds : 6.
5. And-all that-beauty, all that-wealth e'er-gave : 34.



CLASSES OF ADJECTIVES 87

7. O blessings on his kindly voice and on his silver hair!
8. Suddenly there was a gentle little tap on the inside of the lid.
9. The thirteen colonies were now free and independent states.
10. A certain man fell among thieves.
11. Small service is true service while it lasts.—WORDSWORTH.
12. All the air a solemn stillness holds.—GRAY.
13. The good old year is with the past.—BRYANT.

EXERCISE 75

Write sentences containing the following words used as adjectives :—

strong	any	many	brittle	prompt
fair	all	curved	every	distant
both	dutiful	little	modern	neither
some	few	much	each	another
clear	brief	certain	other	several

Figure 1. Six moments at which Thomas Gray's *Elegy Written in a Country Churchyard* (1751) is quoted by a work in the PPA. Top left : *Sheridan's and Henderson's practical method of reading and reciting English poetry* (1796, 3). Top right : *The music of nature* [. . .] (1856, 51). Middle left : *Modern English literature* (1857, 231). Middle right : *Theology in the English poets* (1880, 22). Bottom left : *On the Scansion of English Poetry* (1898, 488). Bottom right : *Two-book course in English* (1906, 87).

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

size, we cannot rely on metadata or known authorial signposts. We need a text-reuse system that can match the fragmented quotations scattered through the archive against a reference body of poetry large enough to capture both the famous and the forgotten.

The Chadwyck-Healey Poetry Corpus

We find that reference corpus in the *Chadwyck-Healey English Poetry Database*. The database was originally conceived in 1990 by Sir Charles Chadwyck-Healey, who, browsing the reference shelves of the British Library, resolved to create "a database of all of English poetry" (Chadwyck-Healey 2020). The core *English Poetry Full-Text Database* was released in 1992, was shipped on five CD-ROMs, and comprised some 165,000 poems by more than 1,250 poets, spanning the eighth century to the start of the twentieth century. The version we work with is substantially larger. Aggregating the original core with later Chadwyck-Healey collections – the second edition, *American Poetry*, *African-American Poetry (1760–1900)*, and *Twentieth-Century English Poetry* – brings our reference corpus to approximately 310,000 individual poems.[3]

The database was conceived from the outset as what Karlin, in his introduction to the second edition, called "a utopian library." Unbound from their original materiality, its texts have "soared into an electronic heaven where they rejoice in an infinity of potential combinations with other texts" (Karlin 1995). For our purposes, the most consequential feature of that utopia is its democratic scope. The database's organizing canon is the *New Cambridge Bibliography of English Literature* (NCBEL), which meant including not just the acknowledged and well-established poetic tradition, but also what a *New Yorker* review described as "the racket" made by "the mute inglorious Miltons" (Lane 1995). This long tail of "ungreat" poets is an important signal for our study: precisely because we do not know in advance which poems the PPA's prosodists reached for, we need a reference corpus wide enough to catch not just the canonical anchors, but also the obscure and the forgotten.

Wide as it is, however, the database has one conspicuous gap: it excludes Shakespeare's plays (which Chadwyck-Healey marketed separately as the *English Verse Drama* database). The *New Yorker* review captured the absurdity well, likening the omission to visiting a liquor store and finding it out of Scotch–"let's face it, out of alcohol" (Lane 1995). To resolve this, we manually supplemented the Chadwyck-Healey corpus with a curated set of additional texts known to be present in PPA works but absent from the Chadwyck-Healey database (first and foremost Shakespeare's dramatic verse, but also several other poems encountered during expert adjudication of manual annotation data).[4] The result is a reference corpus that accounts for what is in all likelihood the single most-quoted author in the PPA, alongside a wider repertoire of texts that the standard Chadwyck-Healey collections do not cover.[5]

[3]Our corpus aggregation follows the procedure described in Heuser (2025), who assembled this combined Chadwyck-Healey corpus for his study of formal patterns in AI-generated verse.

[4]Texts in the supplementary corpus were sourced from scholarly editions where possible (including the Folger Shakespeare and the Eighteenth Century Poetry Archive) and from reputable public-facing repositories (including the Poetry Foundation, All Poetry, and Project Gutenberg). The set was curated by Mary Naydan over the course of work on the dataset.

[5]An account of the editorial decisions governing inclusion and exclusion in the database can be found in Chadwyck-Healey (2020), in Karlin's introductory note to the second edition bibliography (1995), and in the critical exchange between Reimer (1996) and Pocock (1996). For an account of Chadwyck-Healey's place in the development of digital poetry corpora, see Martin (2025).

Alignment, filtering, and binding

Running *passim* over the full PPA and against a subset of the Chadwyck-Healey reference texts[6] (plus our supplementary additions) – 246,724 reference poems in all – yields 1,263,587 local alignment records in the raw export. Each record ties a character span on a PPA page to a span in a reference poem.[7] These alignments resolve to 509,030 distinct poem appearances – (poem, PPA host work) pairs in which a given poem is matched at least once inside a given host – across 7,061 PPA works. Throughout the article, we analyze reuse at the level of the *poem appearance* – a poem's presence in a given host work – rather than the individual alignment record.[8] Running *passim* does not yet result in an analytical corpus. Instead, it results in a set of reuse events that must be filtered on both the PPA "host" side (i.e., which PPA works do the borrowing and therefore "host" a poem?) *and* the poem side (i.e., which poems are admissible to be found?) before it can support any historical claims. Figure 2 traces these successive filtering steps.

On the host side, we restrict the PPA to works tagged as belonging to its *Literary* or *Linguistic* collections. The PPA is divided into six curated collections that reflect the broad scope of historical prosody as a discipline, ranging from dictionaries and rhyming guides to works that use specialized metrical notation.[9] The *Literary* collection gathers versification manuals, handbooks, and treatises, while the *Linguistic* collection gathers grammar books and guides to rhetoric, composition, and elocution. Together they cover the host works in which most prosodic instruction – and most poetic quotation – takes place.[10] Restricting to these two collections leaves 474,298 poem appearances across 6,785

[6]This is a reference-side reduction applied *before* alignment, distinct from the host- and poem-side filtering described below. Because the Chadwyck-Healey corpus contains duplicate texts and poems published after the PPA works that could quote them, we restricted the reference set to a deduplicated subset.

[7]Throughout, "poem" denotes a separately indexed text in the reference corpus. For texts drawn from Chadwyck-Healey, we retain the database's own divisions: an entry may be a whole work or a separately indexed book or section within it.

[8]We adopt the poem appearance as our analytic unit because every question we pursue here concerns a poem's presence *across* host works, not the intensity of its quotation within any single work. This work-level unit carries a cost: a poem appearance registers a poem once per host work however many times that host quotes it, so repeated quotation *within* a work is not counted. We accept this because our lines of inquiry operate at the work level. The choice also absorbs a source of noise on the technical side. Even though *passim* accommodates some textual variation, a single human-visible quotation often varies enough from the reference text – through OCR noise, paraphrase, or partial quotation – that the aligner splits it into several records, so raw record counts exceed the number of visually distinct quotations a reader would count by hand. We therefore reserve *alignment record* for that technical unit. The raw alignments are the product of a parameter sweep. *Passim*'s default settings assume relatively long documentary reuse; verse quotations in manuals are often short, so we reduced n-gram length, minimum aligned span, and gap tolerance, and capped maximum document frequency, then chose final values against a manually annotated gold set. Reported settings: n-gram 15 (default 25), minimum alignment length 25 (default 50), gap 300 (default 600), and max DF 10,000 (default 100). The dataset documentation describes the use of expert-adjudicated manual annotations to construct a page-level evaluation set and select *passim* parameters suited to short poetic excerpts (Thompson et al. 2026).

[9]Collection definitions from https://prosody.princeton.edu/collections/, accessed May 2026. PPA works can belong to more than one collection, so the six collections can overlap (Koeser 2020).

[10]There is significant poetic quotation in dictionaries, but the PPA does not have a representative corpus of dictionaries.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

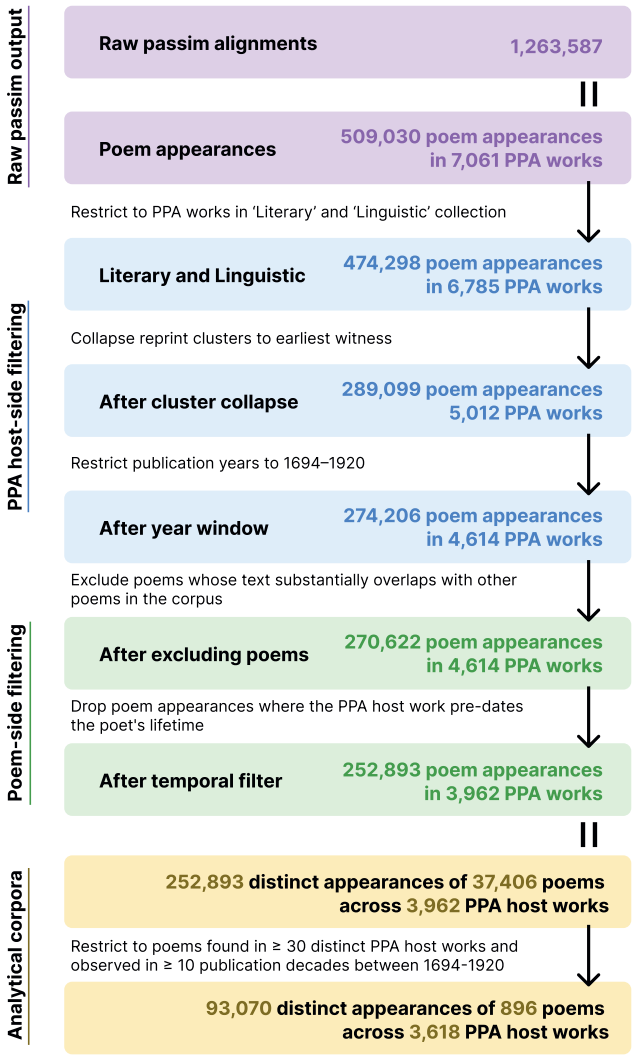


Figure 2. Pipeline from raw *passim* alignments to poem-side temporal filter, with record counts at each stage.

host works. We then collapse reprint clusters to their earliest dated witness (Hostetler, Naydan, and Koeser 2023), so that the same edition appearing in multiple digitization sources is not counted twice. This step brings the dataset to 289,099 poem appearances across 5,012 host works. We also restrict the host record to PPA works published between 1694 and 1920. The lower bound is the first year in which the PPA contains at least five works (anything earlier sits on too thin a host record to support claims); the upper bound is close to the corpus horizon, but set back to avoid edge-of-corpus artifacts where the last few years of digitized works thin out before the corpus formally ends. After the year-window restriction, the host side contains 274,206 poem appearances across 4,614 host works.

Two filters at this stage cannot be derived from the corpus metadata as it ships. The first targets a small group of poems whose textual character makes them unreliable as reference targets; we identify these by hand.[11] The second filter targets biographical plausibility: a PPA work published in 1720 cannot quote a poem by a poet born in 1770. Screening for this requires reliable dates of birth and death for every poet in the reference corpus, and the metadata that ships with Chadwyck-Healey does not support that screening directly. Birth and death years are inconsistently formatted, frequently incomplete, and in some cases simply absent. Using OpenRefine (Delpeuch et al. 2026), we cleaned the Chadwyck-Healey metadata and reconciled it against Wikidata, supplementing and correcting each poet record – where possible – with date of birth (`P569`), date of death (`P570`), and a handful of additional properties.[12] When does a poet's work become available for quotation? For filtering purposes, we treat eighteen as the earliest viable age at which a poet's verse could enter circulation – that is, no PPA work published before a poet's eighteenth year is treated as a credible quotation of that poet, regardless of what *passim* found.[13] The temporal filter leaves an analytical corpus of 252,893 poem appearances of 37,406 poems across 3,962 PPA host works (Figure 2). This is the corpus we use in the "Holding the archive constant" section, where we describe the corpus-wide rate of poetry quotation. In the "Poetry's trajectories" section, we apply a further restriction, keeping only poems that appear in at least 30 distinct PPA works and are observed in at least 10 publication decades. This second analytical corpus contains 93,070 distinct appearances of 896 poems across 3,618 PPA host works. We explain the rationale for the sustained-presence threshold when we introduce the model.

What does the temporal filter actually remove? The largest single source – Henry Baker's 1737 anthology *Medulla Poetarum Romanorum* – accounts for 320 *passim* alignment records, all of them found in PPA works published before Baker's eighteenth year. The pattern reflects a structural mismatch between catalog and corpus: the *Medulla* is recorded as a single Chadwyck-Healey poem, but its content is made up of translated extracts from Latin poets who are centuries older. The alignments found by *passim* are real; the attributions are not. Similar ambiguities arise in works that incorporate earlier texts. Longfellow's *Christus: A Mystery* (1872) is built substantially from scriptural paraphrase and passages that elocution manuals like Blair's *Lectures on Rhetoric* (1783) and Sheridan's *Lectures on the Art of Reading* (1775) had been quoting compulsively for a century before Longfellow set them in dramatic verse. The alignment detects biblical commonplaces

[11] These fall into a few types. Cento poems – compositions made up of lines borrowed from other poets and arranged into a new work – are the most obvious case: Laman Blanchard's *Ode to the Human Heart* (1842) is one such example (McCutcheon 2012). Any *passim* match against its text is really a match against the source poems it stitches together. Adaptations and free renderings present a related problem. Richard Jago's *Adam* (1784), an oratorio compiled from *Paradise Lost*, shows up in alignments because its text shadows Milton's so closely that the alignment is in effect against Milton with noise. Removing these cases takes the count from 274,206 to 270,622 poem appearances, across the same 4,614 host works.

[12] Sex or gender (`P21`), occupation (`P106`), and country of citizenship (`P27`). The cleaned and reconciled poet metadata is released alongside this article: 10.5281/zenodo.20597001.

[13] Where biographical anchors exist, we apply them as strict rules; where they are partial, we fall back through a short cascade of alternatives, with the death year as a last resort. When no biographical anchor exists at all (in most cases, this means when a poem is attributed to an anonymous poet, or to a poet whom neither Wikidata nor Chadwyck-Healey can reliably date), we keep the excerpt rather than discarding it. Anonymous and untraceable poetry is a feature of the historical record, and dropping it for the convenience of a clean filter would erase a layer of the reception history we are trying to recover. We also deliberately do not use edition publication years as biographical anchors (even though they are available for many poems). A poem may have circulated long before the edition represented in the database, so using that edition's date as a cutoff could exclude genuine earlier quotations.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

that the rhetorical tradition had consecrated; it is not detecting an anachronistic reading of Longfellow in 1780.[14] Pound's *Cantos* (published in installments between 1917 and 1969) presents an even more telling case. The phrase "the evil that men do lives after them" is found in roughly 30 PPA works,[15] but the corresponding passage in *Canto LXXX* reads:

> The evil that men do lives after them'
> well, that is from Julius Caesar
> unless memory trick me
> who crossed the Rubicon up near Rimini

Pound is not being quoted by the elocutionists; he is quoting Shakespeare, even naming the source, and turning the act of citation itself into poetic material. The biographical filter discards the alignment correctly (Pound was born in 1885, so the cited manuals all precede his eighteenth year) but the deeper point is that the *Canto*'s passage and the rhetorical tradition share a single Shakespearean ancestor, not a line of transmission. What *passim* has caught in cases like these is the shared language of a literary culture: scripture, Shakespeare, the Latin poets in translation. The temporal filter reduces the risk of silently misattributing that language to a later poet.

The shape of the reception record

The 1694–1920 analytical window is where will spend most of our interpretive time, but before we do so we want to give a picture of the full dated PPA record. Among poems whose first appearance in the PPA can be dated between 1532 and 1929, the picture of poetic quotation is one of extreme concentration. Most poems enter the record only lightly: 16,378 poems (42 percent of the population) are quoted in only *one* PPA work across the entire 1532–1929 window. At the other end of the curve, a small core of canonical anchors carries a disproportionate share of the evidence: the top 15 poems alone, listed in Table 1, account for nearly 5 percent of *all* appearances.

The top of Table 1 reads as a list of the usual suspects. Milton, Shakespeare, and Pope between them account for eight of the top 15 rows; the King James Bible occupies three of them; Byron, Longfellow, and Gray fill out the rest. What kind of distribution produces this shape, and what kind of process sustains it? The PPA's "quotation economy" resembles what economists since Rosen (1981) have called an "economy of superstars" – a market in which small differences in initial visibility translate into very large differences in eventual reach, sustained by the reproducibility of whatever is being consumed. Once a PPA work quotes *Hamlet* (e.g., to illustrate a point about iambic pentameter), that line "costs" nothing to copy into the next manual. In other words, once *Hamlet* has become the canonical illustration, reaching for it is easier than searching out an unused alternative.

We can also put this imbalance into a single number. The Gini coefficient, introduced by Gini (1912) to measure inequality, tells us how unevenly appearances are distributed across poems. A value of zero would mean that every poem appears in the same number of PPA host works; a value of one would mean that all appearances belong to a single poem. In the full dated PPA record, across 38,914 poems and 267,237 appearances, the Gini coefficient is 0.72.[16] The number confirms what Figure 3 shows visually: quotation in the PPA is highly concentrated. A small group of poems is used again and again, while most poems appear only lightly. The 16,378 poems quoted in a single PPA host work – such as William Cowper's *On the High Price of Fish* (1803), John D. M'Kinnon's *The Mohawk* (1802), and Christina Rossetti's *Seldom "can't"* (1872) – make up 42 percent of the distinct poems in the corpus, but only 6 percent of all quotation events.

The rank view, however, flattens time. Knowing that *Paradise Lost* sits at the top across the entire archive tells us something about its total mass in the record, but nothing about its *momentum* – whether it gained or lost ground after its first appearance in 1679, or whether the population of poems below it is growing, shrinking, or churning. Concentration of the kind we have just described could in principle be produced in several ways: by a stable core of canonical anchors that have been quoted at a steady rate for two centuries, or by a much narrower historical regime – for example, a late-Victorian boom in pedagogical publication that mechanically inflates the head of the distribution at the expense of everything around it. To tell these stories apart, we look at the record as it unfolded over time.

Figure 4 plots, for each PPA host publication year between 1532 and 1929, the number of distinct poems quoted at least once in PPA works dated to that year (blue), alongside the number of distinct poets to whom those poems are attributed (orange). Both series are smoothed with a ten-year rolling mean to quiet year-to-year noise. A gray ribbon beneath the main panel reports total host-work pages per year, a reminder that the archive itself is changing across the same window. Before about 1700, the archive is relatively thin. From then on it thickens, slowly through the eighteenth century and more steeply through the nineteenth century. That changing thickness is worth keeping in view alongside the two curves above it, because it sets the conditions under which any quotation can happen at all. A year for which the PPA holds only a few hundred pages offers fewer occasions on which a poem could be quoted than a year that holds tens of thousands.

Against that backdrop, the curves tell a story of two regimes. Through the seventeenth century and the early eighteenth century, the counts of distinct poems and distinct poets run close together near the floor; only a few dozen poems are quoted in a typical year, drawn from a similarly thin pool of poets. A first sharp jump arrives in the late 1770s and again in the 1790s, when a relatively small set of compendious PPA works briefly lifts the count to between 700 and 1,000 distinct poems in a single year. Largely, what we are seeing here is the effect of multi-volume editions of the English poets and the literary histories that proliferated after *Donaldson v. Becket* (1774) (Bonnell 2008; St. Clair 2004). That decision struck down the London booksellers' claim to perpetual copyright and pulled a large body of older verse into a clearly defined public domain, where competing publishers could now reprint it in bulk.[17]

[14] The filter retains Longfellow matches from 1825, the year he turned eighteen. *Christus* incorporates earlier works by Longfellow, including his 1851 poem *The Golden Legend*, as well as biblical passages. For later hosts, the date filter alone cannot tell whether shared biblical wording was taken from Longfellow's work or quoted from scripture independently.

[15] Among others, in Scott's *Lessons in Elocution* (1814), Vandenhoff's *The Art of Elocution* (1847), and Clark's *Practical Rhetoric* (1886).

[16] For comparison, recent global citation inequality at the author level in modern science has been measured at Gini values rising from 0.65 to 0.70 over 2000–2015 (Nielsen and Andersen 2021).

[17] The bumps we notice in Figure 4 at the end of the eighteenth century are dominated by Warton's *History of English Poetry* (1774–1781); *The Works of the English Poets*, commissioned by a group of London booksellers (the consortium) to publish a uniform edition of English poetry; Robert Anderson's *Complete Edition of the Poets of Great Britain* (1792–1795); and volumes of John Bell's *Poets of Great Britain* (1777–1783). Each of these volumes carries hundreds of distinct quotable poems, so a single year's worth of new releases can dominate that year's distinct-poem count.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

Table 1. Most and least frequently hosted poems in the full dated PPA record

Rank	Work title	Author	Appearances	First	Last
1	Paradise Lost	John Milton	1,326	1679	1929
2	Hamlet	William Shakespeare	1,287	1702	1929
3	Psalms	King James Bible	1,145	1644	1928
4	Julius Caesar	William Shakespeare	909	1702	1929
5	Macbeth	William Shakespeare	877	1702	1929
6	The Merchant of Venice	William Shakespeare	844	1718	1928
7	An Essay On Criticism	Alexander Pope	833	1712	1929
8	Proverbs	King James Bible	757	1644	1927
9	Elegy Written in a Country Churchyard	Thomas Gray	735	1756	1929
10	Job	King James Bible	698	1644	1927
11	Don Juan	Lord Byron	694	1810	1929
12	Christus: A Mystery	Henry Wadsworth Longfellow	687	1825	1929
13	Othello	William Shakespeare	680	1702	1929
14	Childe Harold's Pilgrimage	Lord Byron	678	1814	1929
15	As You Like It	William Shakespeare	655	1707	1929
		...			
38910	On The High Price Of Fish	William Cowper	1	1923	1923
38911	The Mohawk	John D. M'Kinnon	1	1909	1909
38912	Seldom "can't"	Christina Georgina Rossetti	1	1915	1915
38913	Greater Memory	Arthur William Edgar O'Shaughnessy	1	1916	1916
38914	To A Cricket	William Cox Bennett	1	1912	1912

Note: "Appearances" counts distinct PPA host works in which the poem is matched at least once; repeated matches inside the same host work count once. "First" and "Last" give the earliest and latest publication dates of host works in which a poem is matched.

The count falls back through the early nineteenth century as that wave passes; from the mid-1840s, it begins a longer and steadier climb, reaching some 2,000 distinct poems quoted per PPA host year by 1900. The sharp drop at the very right edge of the plot reflects the PPA's coverage horizon rather than any change in the practice.

Figure 4 is useful for scale and timing, but the count it reports is a relatively naive tally: each point records how many different poems and poets appear in PPA works published that year, without any sense of how the figures relate to the actual *volume* of writing dated to the same year. Years in which the archive adds large host works – anthologies, literary histories, and elocution readers – will register as years in which many more poems were quoted, even when the underlying instructional habit barely shifts. The next section asks what remains of the rising curve once that confound is accounted for.

Holding the archive constant

The trouble with a year-by-year count of distinct poems is that it cannot distinguish between two situations: the practice of quoting verse itself may have been expanding – prosodists in 1850 reaching for more poetry than their predecessors in 1750 – or the practice may have actually been stable, and the rising count may reflect nothing more than the fact that there were more host works in which a poem could be quoted at all (i.e., more pages of opportunity, an increase the gray ribbon makes visible). The way to tell these stories apart is to ask, for each year, not how many distinct poems were quoted, but how many were quoted per page of available host text. This gives us a common scale for comparing years with very different amounts of archived writing.

The simplest version of the measure is to take the count of distinct poems in year y, call it $A(y)$, for the size of the *active* repertoire that year, and divide it by the total number of host-work pages dated to that year, $P(y)$.[18] The result, $A(y)/P(y)$, is a rate: events per unit of opportunity. In our setting, the events are distinct poems appearing in the PPA corpus and the opportunity is the pages on which they could have appeared. A rising rate means that more distinct poems are represented per page of archived host text. Simply plotting that ratio year by year, however, would produce a curve too jagged to read. Some years carry only a few hundred pages of host writing, and in those years, a single dense anthology can push the ratio to extreme values; other publication years carry tens of thousands of pages spread across many host works and produce a much steadier figure. The signal we want is that of the underlying *trend*, which sits buried inside this year-to-year noise.

[18]Pages provide an approximate measure of textual exposure. Page size, typography, layout, and the balance of prose and verse can vary across works and periods. The offset adjusts for recorded page volume, but does not fully standardize the amount or kind of text that volume contains.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

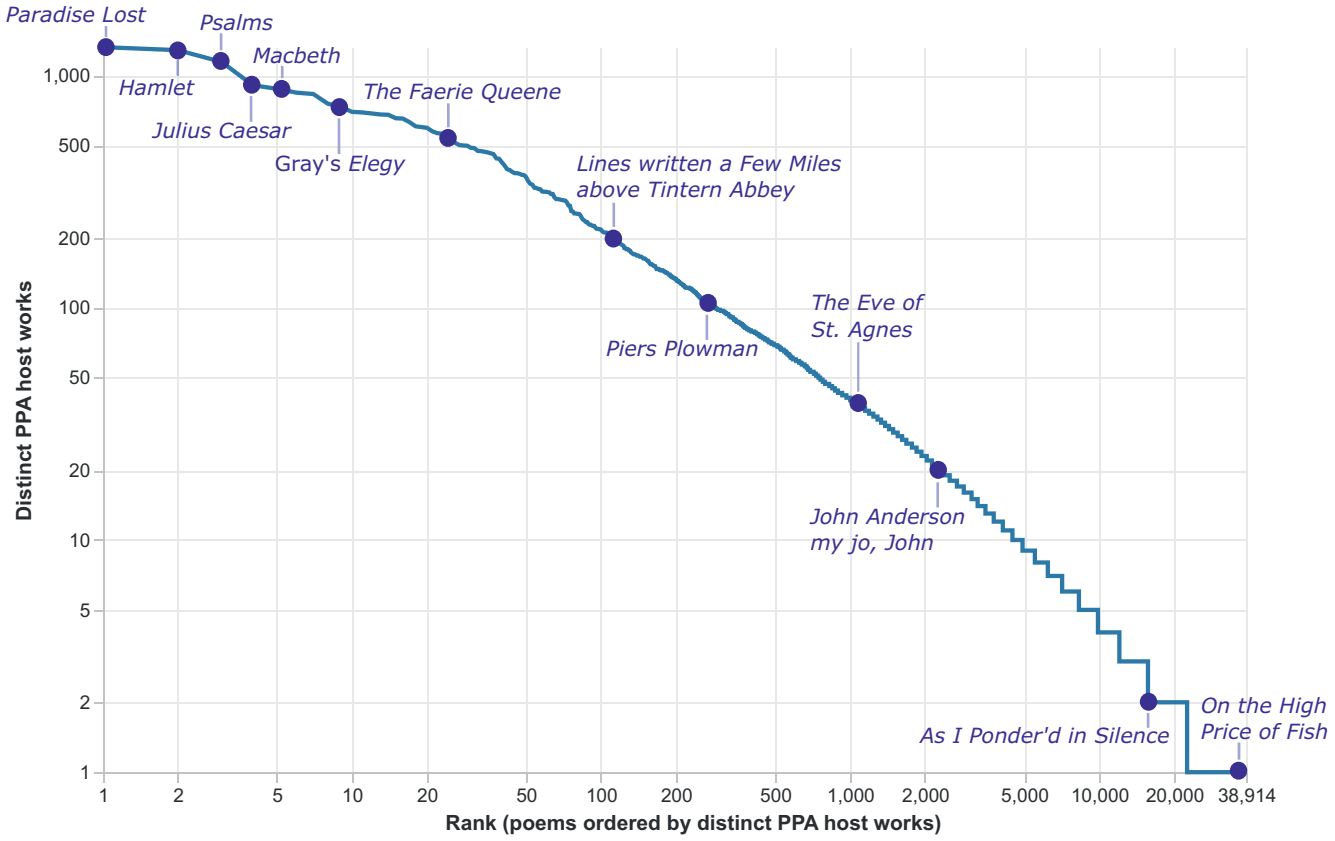


Figure 3. The number of distinct PPA host works quoting each poem, plotted against the poem's rank, on log–log axes, for all 38,914 poems in the full dated PPA record. The line is approximately straight across most of its range, the visual signature of a heavy-tailed distribution: a small core of poems is quoted very widely, while the great majority appear in only a handful of host works. We have manually annotated a selection of poems along the curve to indicate where familiar works sit relative to the broader population.

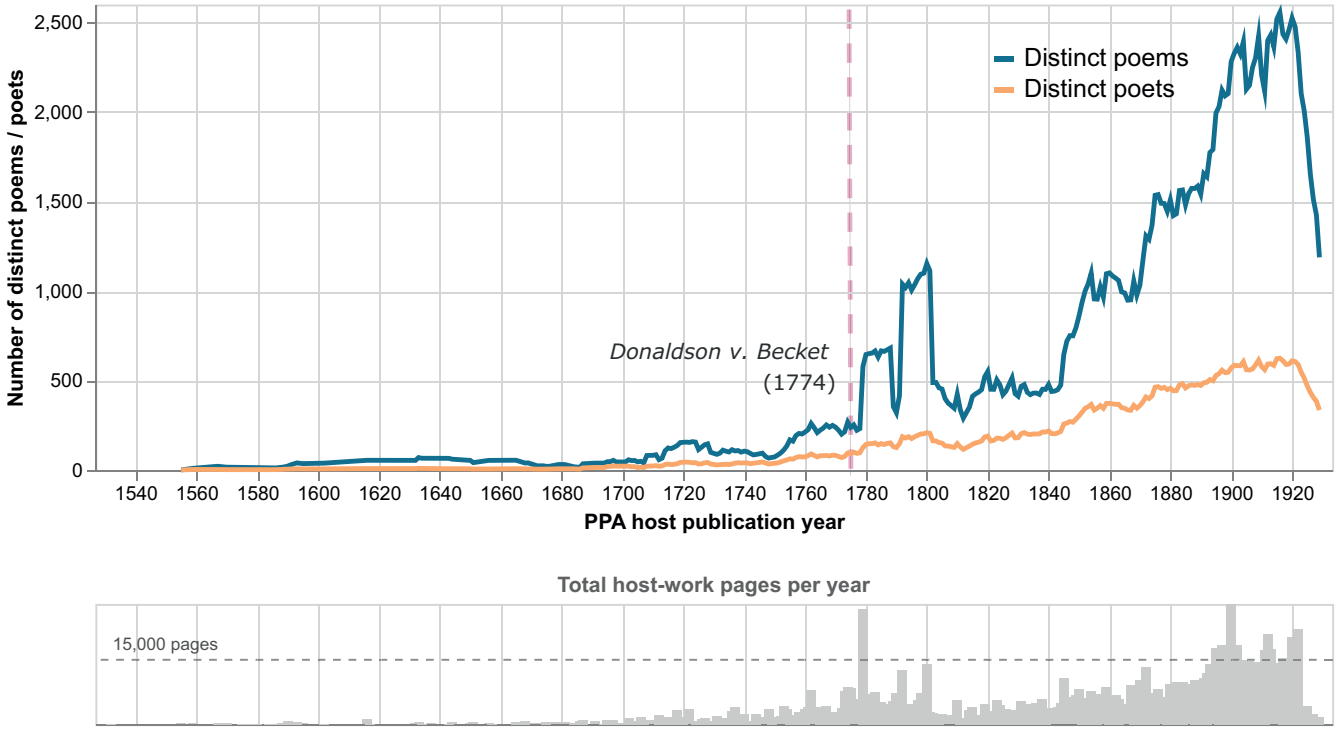


Figure 4. Distinct poems and distinct poets quoted in PPA host works dated to each publication year, 1532–1929 (10-year rolling mean). The gray ribbon below the main panel reports total host-work pages per year, with a dashed reference at 15,000 pages.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

To recover it, we have to fit a model that lets us separate the trend from the fluctuation. In short, the model has to do three things:

- it has to express the rate, $A(y)/P(y)$, rather than the raw count $A(y)$;
- it has to let that rate rise and fall smoothly with time, since we have no reason to believe the practice was constant;
- it has to accommodate the fact that the year-to-year counts are highly variable in a way that simple count models cannot handle.

We discuss these one at a time. Although the estimand is a rate, the observed outcome we model year by year is the count $A(y)$; the rate enters through an offset on page volume. An offset is a term added to the right-hand side whose coefficient is fixed at one, so the model still predicts counts, but its parameters describe the poems-per-page rate and how that rate changes over time. The natural model for a count outcome is the Poisson family, which puts a logarithm on the *expected* count:

$$\log \mathbb{E}[A(y)] = \alpha, \tag{1}$$

where α is a baseline log-count, the average level the count tends toward across the window. To turn this into a model of the rate rather than the count, we add $\log P(y)$ to the right-hand side and fix its coefficient at one. By forcing the coefficient to one, we tell the model that the relationship between pages and poems is structural – if the available host writing doubles, we assume that the count of quoted poems should roughly double with it – rather than something the model has to discover from the data[19]:

$$\log \mathbb{E}[A(y)] = \alpha + \log P(y). \tag{2}$$

What this equation gives us is an estimate of how many different poems are quoted per page of host writing. Subtracting $\log P(y)$ from both sides and exponentiating gives $\mathbb{E}[A(y)]/P(y) = e^{\alpha}$, the rate of distinct poems quoted per page of host writing. The page-volume term has, in effect, been pulled out of the response and absorbed into the denominator; the intercept α now stands for the log of the per-page rate, which is the quantity we set out to measure.

The second requirement is that the rate be allowed to change with time. As written, the model above assumes a single rate throughout the window – it gives us one overall level, but cannot show us *how* the rate changes across two centuries. What we want to recover, however, is something whose shape we do not yet know: the model should be free to discover the historical trajectory from the data. We therefore augment the intercept with a smooth function of year:

$$\log \mathbb{E}[A(y)] = \alpha + s(y) + \log P(y). \tag{3}$$

Here, α is an intercept and $s(y)$ is a natural cubic spline. Together they describe the log-rate as a flexible curve fitted to the data, free to bend where the evidence calls for it to bend.[20] The number of bends the spline is allowed to take is a tuning choice: we fit the model at several candidate complexities and keep the one with the best statistical fit.[21]

The third requirement concerns how the model represents the year-to-year variability in the data – the inevitable wobble of any count series around its underlying trend. In statistical terms, this is the choice of *likelihood*, the assumption the model makes about how the observed counts are distributed around the values it predicts. Different likelihoods make different assumptions about how variable the data can be. The Poisson family, which we have been working with implicitly so far, assumes that the variance of the response equals its mean.[22] In our case, this assumption is violated. The year-to-year variance of $A(y)$ is several times its mean: a single dense anthology landing in one publication year can pull the count far away from the smooth trend the spline is trying to follow, and the Poisson has no room to accommodate that kind of departure. We therefore use a more permissive likelihood, the Negative Binomial, which adds a single parameter – call it α_{disp}, for dispersion – that allows the variance to grow faster than the mean. Conceptually, the Negative Binomial says: the count fluctuates around the model's prediction, but it is allowed to fluctuate more than a strict count-distribution would predict. How much more is itself fitted from the data. The full model can now be written down compactly:

$$A(y) \mid \mu_y \sim \text{NegBin}(\mu_y, \alpha_{\text{disp}}) \tag{4}$$

$$\log \mu_y = \alpha + s(y) + \log P(y). \tag{5}$$

Equation (4) says that the count of distinct poems in a given year is a draw from a Negative Binomial distribution with mean μ_y and a dispersion α_{disp} controlling how widely the actual counts may scatter around that mean. Equation (5) says that the mean itself is determined by three things: the intercept α, the smooth time trend $s(y)$, and the page-volume offset $\log P(y)$ that holds the corpus thickness constant.[23]

One last decision concerns the window across which we fit. We restrict the model to the years 1694–1920. We treat 1694 as the start of the modeled window because it is the earliest year in which at least five PPA host works appear in our corpus, a practical starting point that avoids the sparsest early years of the archive.[24] At the other end, 1920 is set by the PPA's coverage horizon, beyond which the archive thins again for reasons of digitization and copyright rather than historical practice.

[19] An ordinary predictor would let the model estimate how strongly the page count influences the poem count. The offset, by fixing the coefficient at one, declares this relationship in advance: more pages of host writing should mean proportionally more poems quoted, and what we want the model to discover are the departures from this proportionality, that is, the moments and decades in which prosodic writing reaches for more poems per page than the corpus's sheer volume would predict.

[20] The natural-cubic choice is standard for smoothing in this kind of setting because it avoids the wild behavior at the edges of the window that an unconstrained polynomial fit would produce.

[21] The selection criterion is the Akaike information criterion (AIC), a standard goodness-of-fit measure that penalizes a model for adding parameters it does not need. In our case, AIC selects a spline with seven degrees of freedom – enough flexibility for the rate to bend across the long span of years the model covers, but not so much that the curve chases noise.

[22] In other words, if the model predicts an expected count of, say, 50 poems in a given year, the Poisson assumes that the actual count will fluctuate around that prediction with a variance also equal to 50 – that is, a standard deviation of about 7. Real-world count data rarely behave that obediently.

[23] We select the dispersion parameter by AIC over a coarse grid of intercept-only Negative Binomial models with the page-volume offset. We then hold it fixed while choosing spline complexity by AIC. In the reported fit, $\alpha_{\text{disp}} = 0.5$ and the post-fit Pearson dispersion diagnostic comes in at 1.01 – essentially one, suggesting that the Negative Binomial is accounting for the extra variability in the counts.

[24] Before this point, the archive thins to one, two, or three host works per year, and a rate computed from such a thin denominator is too volatile to support stable estimation. A single early-modern host work that quotes a thousand poems across 500 pages would produce a per-page rate that swamps everything around it.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

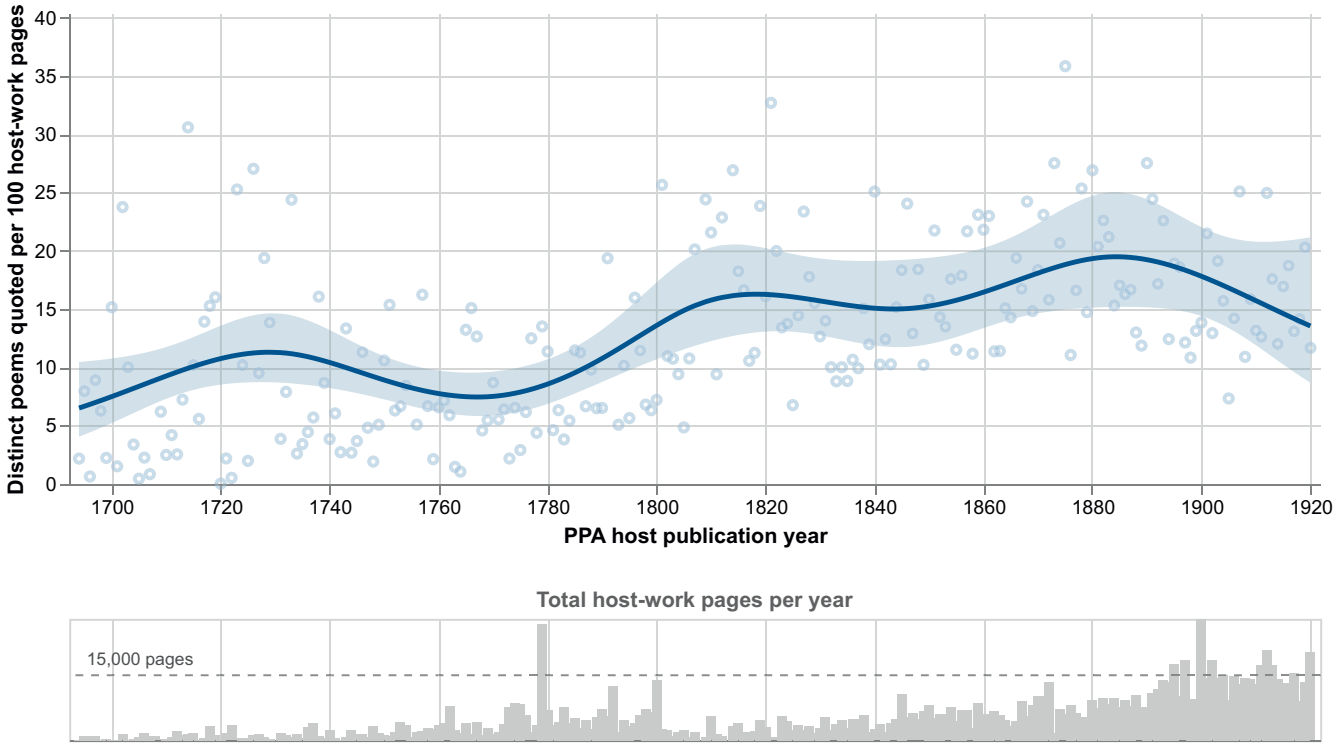


Figure 5. Distinct poems quoted per hundred pages of writing about verse, for each year between 1694 and 1920. The solid line is the fitted rate from a Negative Binomial model with a natural cubic spline on year and a log-page-volume offset; the shaded band is the 95 percent confidence interval; the blue points are the raw annual rates. The ribbon below reports total host-work pages per year.

Figure 5 shows the result. The fitted rate is plotted as a solid line, with a 95 percent confidence interval around it; the raw per-year ratios are plotted as faint dots. The eighteenth century holds relatively steady. From 1700 through the 1780s, the rate sits between 7 and 11 distinct poems per hundred pages of host writing – roughly one quoted poem for every 9 to 13 pages of instructional prose. That low and stable floor is consistent with what we know from individual pedagogical texts of the period: William Enfield's *Speaker* of 1774, perhaps the most widely used elocution anthology of the late eighteenth century, drew 75 of its 113 oratorical extracts from Shakespeare alone (Salamone 2017). A pedagogy organized around so narrow a base of authoritative passages would not be expected to produce a high per-page rate, and our figure confirms that it did not.

From the 1780s, the curve begins to climb. The rise is at first gradual, reaching about 16 poems per hundred pages by 1820 and then holding at roughly that level through the 1830s and into the 1840s. This first phase of the climb falls in the decades we have already discussed in connection with the multi-volume reprint editions that followed *Donaldson v. Becket*: the available reservoir of English poetry was widening, and pedagogical and prosodic writing were beginning to reach into it. From the late 1860s onward, however, the rate makes a second push upward, peaking near 20 poems per hundred pages around 1880 before softening slightly through the rest of the modeled window. That second acceleration coincides with a set of changes in the institutional setting of pedagogical writing itself: in 1867, English literature was introduced as an optional special subject in British schools, and in 1870, Forster's Education Act established school boards to expand elementary education, with English literature recognized the following year as a class subject in the upper grades (Martin 2012). The kind of writing the PPA captures after 1870 was being produced for, and in response to, a vastly expanded teaching apparatus. That the per-page rate moves upward in those decades suggests that the practice was widening not only because there was more poetry to draw on, but because more pedagogical writing was being produced for classrooms that asked it to draw from a wider pool.

Poetry's trajectories

The model in our previous section treats quotation as a single, undifferentiated practice and smooths over the movements of specific poems that compose it. A natural extension of the model presented above is to take that trendline apart and look at *which* poems were being reached for while others were perhaps being quietly set aside. To see those individual movements, we need a different kind of model: one that gives every poem its own trajectory – its own rising or falling line across the centuries – while pooling information across the corpus, so that poems with thinner evidence behind them can be read against the better-attested ones. Pope's *Essay on Criticism* and Wordsworth's *Tintern Abbey* make this analytically possible in different ways: both have decades of repeated appearance in prosodic writing, with enough quotations distributed across enough host works that a curve can be fit through them. The object of study is the population of poems that look like these – poems with what we will call *sustained presence*, reached for repeatedly by PPA authors across multiple decades of their available life in print. Some of these poems are familiar canonical anchors quoted across the entire 1694–1920 window; others are less obvious, with concentrated reception in particular decades, or with a quiet persistence that registers across

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

two centuries without ever rising to prominence. What changes within this population – which sustained-presence poems rose, which fell, and which held steady – is what the model is built to recover.

To model how a poem's presence in the PPA changes over time, we need two things from the evidence: a meaningful number of host works that quote the poem, and those host works distributed across enough decades that we can actually see change. We require, for inclusion in the model, at least 30 distinct PPA host works quoting a poem somewhere within the 1694–1920 modeled window, with those host works spread across at least 10 distinct decades.[25] The resulting modeled population contains 896 poems – a long way from the 37,406 distinct poems in the full found-poems corpus, but considerably wider than the small set of canonical anchors that dominate the rank distribution discussed in "The shape of the reception record" section.[26]

To recover the per-poem movements that the model described in the "Holding the archive constant" section smooths over, we need something more flexible. Every poem now needs its *own* trajectory across the decades in which it was eligible to be quoted, and these individual trajectories need to live alongside a shared backdrop: a corpus-wide tendency that describes the practice as a whole and that each poem either follows, exceeds, or falls behind. The model that does both at once is a hierarchical one: it estimates the shared trend *and* the poem-specific deviations simultaneously. Those two estimates then inform each other through pooling: evidence about how poems behave on average partly stabilizes estimates for any individual poem. A poem with less attestation in a particular decade has its trajectory partly pulled toward the shared trend; a poem with dense attestation pulls the shared trend in its direction. Where a poem is well attested, its fitted trajectory depends more strongly on its own record; where evidence is sparse, information from other poems has more influence. The fitted trajectories therefore read each poem against the corpus, with the shared trend serving as the regularizing context against which individual poems are interpreted.

The quantity the model is built to fit, at the level of a single poem in a single decade, is the share of prosodic writing that quoted the poem. For each eligible poem i and each decade t in its eligibility window, we capture two numbers. The first, k_{it}, is the number of distinct PPA host works published in decade t that quote poem i at least once. The second, n_t, is the total number of distinct PPA host works in scope in decade t. The ratio k_{it}/n_t is the prevalence of poem i in decade t: the share of prosodic writing in that decade that engaged the poem at all. A prevalence of 0.10 in the 1850s, for instance, means that 1 in 10 PPA host works published in that decade quoted the poem; a prevalence of 0.40 means 4 in 10. This quantity – one number per poem per decade across each poem's eligibility window – is what the model takes as input and what its fitted trajectories ultimately describe.

The natural starting model for a count out of a known total is the binomial:

$$k_{it} \mid n_t, p_{it} \sim \text{Binomial}(n_t, p_{it}), \tag{6}$$

where p_{it} is the probability that any given PPA host work in decade t quotes poem i. As in the "Holding the archive constant" section, however, the simple count distribution understates the variability we actually observe in the data. We therefore use the beta-binomial instead, which adds a single parameter κ that allows the variance to grow beyond what the binomial would permit.[27] The mean of the beta-binomial is still $n_t \cdot p_{it}$; what changes is how tightly the actual counts are required to cluster around that mean.

What remains is to say what determines p_{it} – each poem's prevalence in each decade. This is where the hierarchical structure of the model enters. We write the log-odds of p_{it} as the sum of three pieces, each describing a different layer of what makes a poem visible in prosodic writing at any given moment:

$$\text{logit}(p_{it}) = \underbrace{\alpha + s(t)}_{\text{shared time curve}} + \underbrace{b_i^0 + b_i^1 \cdot t}_{\text{poem-level deviation}} + \underbrace{u_{a(i)}}_{\text{author-level shift}} . \tag{7}$$

The first piece, $\alpha + s(t)$, is the shared time curve: a corpus-wide tendency that captures whatever prosodic writing in the sustained-presence population was doing on average at a particular time. Every poem in the model inherits the shape of this curve as its baseline trajectory. The shared curve is what the average poem looked like across the modeled window. Holding the shape of that average in view is what makes the per-poem trajectories interpretable as deviations from it: a poem whose fitted trajectory carries a positive slope was not necessarily rising in absolute terms across its eligibility window, but outpacing the average poem; a poem with a negative slope was not necessarily fading from the practice in absolute terms, but lagging behind it.

The second piece, $b_i^0 + b_i^1 \cdot t$, is the poem-level deviation: this is the part in the model that gives each poem its own trajectory. The intercept b_i^0 captures whether poem i tends to sit above or below the corpus average across its eligibility window. The slope b_i^1 captures whether the poem is gaining or losing prevalence in prosodic writing (on top of whatever the shared trend is doing in the same decades). A poem with a positive slope was gaining ground relative to the shared corpus trend; a poem with a negative slope was losing ground within the quotation repertoire, falling behind the shared corpus trend. These two parameters are drawn jointly from a prior distribution, which is what allows the model to pool information across poems. A canonical poem quoted in

[25] Thresholds of this kind always involve a trade-off. A lower bar would let more poems into the model but on evidence too thin to be reliable; a higher bar would tighten reliability at the cost of dropping interesting cases – particularly poems by later-born poets whose eligibility windows are short against the corpus horizon of 1920.

[26] Importantly, for each of these poems, we further restrict its time series to the decades in which it could plausibly have been quoted at all, on the same biographical grounds as the temporal-plausibility filter described in the "Alignment, filtering, and binding" section. The rule is the same priority cascade: a poem's eligibility window begins eighteen years after the poet's birth. In survival analysis, this construct is conventionally called the "at-risk" window, the period in which a unit is exposed to the possibility of the event under study. We use "eligibility" instead because the survival-analysis term carries epidemiological connotations (the event being typically adverse) that sit awkwardly when the event in question is being quoted in a prosody manual. The filter, described in the "Alignment, filtering, and binding" section, applied to alignments, removed implausible *passim* matches before they could enter the analytical corpus. The restriction here is different in kind: it concerns how the model should read the *absence* of quotation in decades when a poem could not yet have been quoted at all. A Tennyson poem shows zero quotations in the 1700s; without left truncation, the model would read that zero as evidence of pedagogical neglect, when in fact it reflects biographical impossibility. Restricting the time series to a poem's eligibility window ensures that the trajectory we fit is about reception, not about structural absence.

[27] The beta-binomial is to the binomial what the Negative Binomial was to the Poisson in the "Holding the archive constant" section: a more permissive likelihood that absorbs the extra variability without forcing the model to misfit the trend. The parameter κ controls how much extra variance the model can accommodate; like α_{disp} in the corpus-wide model, κ is fitted from the data.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

hundreds of host works will have its intercept and slope estimated almost entirely from its own data points. A poem with sparse evidence will have its estimates partly stabilized by the corpus-wide distribution of intercepts and slopes, so that its trajectory is read as a plausible variation on what other poems are doing rather than as a wild swing supported by limited data.

The third piece, $u_{a(i)}$, is an *author-level shift*: an adjustment, shared across all poems by the same author a, that allows the model to account for authors whose works tend systematically to be quoted more or less than the average across the corpus. Some authors contribute many poems to the modeled population – for example, Shakespeare contributes 37 plays alongside his sonnets, while Thomas Gray's presence rests almost entirely on his *Elegy* – and a poem's prominence in the PPA may reflect, in part, the standing of its author. Treating each of an author's works as a fully independent observation would overlook this shared tendency. The author shift captures that tendency as a constant adjustment across the modeled window. Each poem still has its own intercept and slope, allowing poems by the same author to differ both in their overall prevalence and in how they gain or lose ground relative to the shared time curve.[28]

Once fitted, the model offers us, for each poem in the sustained-presence population, not a single trajectory but a distribution of trajectories that the evidence could support – all the plausible ways that poem's life in the PPA might have unfolded given what we observe in the PPA.[29] From this distribution, we can read two things about each poem. The first is the poem's overall standing relative to its peers across the modeled window – whether it sat above or below what prosody writing was doing on average – captured by the intercept b_i^0. The second is the direction in which the poem moved through the window relative to the shared trend, captured by the slope b_i^1: a positive slope means the poem was gaining ground relative to the shared trend, while a negative slope means it was losing ground relative to that trend. The width of the distribution around each of these tells us how confident we can be: tighter for canonical anchors with abundant evidence, wider for poems whose record is thinner.

Risers, decliners, and anchors

The fitted model lets us ask which poems gained ground relative to the shared trend over the long nineteenth century, and which lost ground. Table 2 lists 15 poems with credibly negative slopes; Table 3 lists 15 with credibly positive slopes. Each entry gives the poem and its author, the posterior-mean slope b_i^1, and the 95 percent credible interval for that slope.

Read side by side, the two lists describe a renewal of pedagogical use as much as a substitution of one canon for another. The decliners (Table 2) are dominated by the verse pedagogy of the Augustan and neoclassical tradition. What fades is the eighteenth-century apparatus that had made ancient poetry available to English readers as instruction: translations, verse-essays, and compiled *loci* through which classical authority was carried into the English schoolroom (Benedict 2015). Virgil is the clearest case. Three of Dryden's translated *Pastorals* (1697) sit among the strongest decliners. Joseph Warton's verse-translation of the *Georgics* (1753) falls with them, and Christopher Pitt's *Æneid* (1740) closes the table. Horace follows the same pattern. The *Art of Poetry* declines twice, once in Roscommon's 1680 version and again in William Duncombe senior's 1757 translation. Roscommon's own *Essay on Translated Verse* (1684) belongs to the same economy of classical mediation, and Waller's commendatory poem on Roscommon's Horace – originally part of the paratext that presented the translation to readers – declines beside it. Dryden's verse-translations of Ovid's *Metamorphoses* (Books I and XII; 1693 and 1700) and the *First Satire of Persius* (1693), together with Henry Baker's *Medulla Poetarum Romanorum* (1737) – a much-used compendium of translated Latin poetic passages – belong inside the same instructional economy. Sheffield's *Essay upon Poetry* (1682) belongs to the verse-critical tradition that helped readers understand poetry as an art governed by doctrine. And it did so in a form – heroic couplets – that revered classical literature as the standing measure of poetic value.

Blackmore's *Prince Arthur* (1695) gives the table of decliners a sharper edge. It came to serve as a negative exemplar within the same neoclassical pedagogy: a failed Whig epic through which readers could recognize false grandeur and epic overreach. Pope helped fix that role. In *Peri Bathous, Or the Art of Sinking in Poetry* (1728) and the *Dunciad* (1728–1743), Blackmore was preserved as "Everlasting Blackmore" and as a poet so dull that he could put lawyers to sleep. His decline, however, was already under way. Even though Blackmore had enjoyed real support – and his *Prince Arthur* was widely read – by the 1720s, his audience was narrowing, and later references often preserved him as a name for poetic failure rather than as a poet to be read (Rounce 2014).

The risers, presented in Table 3, give this change its nineteenth-century center of gravity. Coleridge's *Rime* (1798) leads the table, Wordsworth appears five times, and Keats, Shelley, and Burns help define a repertoire built around inward feeling and narrative imagination. Across the eighteenth century, verse translation and verse criticism had supplied many of the examples through which poetry was taught. By the late nineteenth century, that work was being done by *Tintern Abbey* (1798), the *Ode on Intimations* (1807), the *Rime* (1798), *Christabel* (1816), and *The Eve of St Agnes* (1820). This movement belongs to a larger revaluation of poetic authority. The "Pope controversy" of 1819–1826 was never only a quarrel over Pope's reputation; it reopened the question of what kinds of verse could still stand near the center of the English canon, and on what grounds (Chandler 1984).[30] The same pressure can be described as a gradual lyricization of poetry: an uneven nineteenth-century

[28] All parameters are given weakly informative priors: Normal distributions for regression coefficients, Half-Normal distributions for the standard deviations of the per-poem and per-author random effects, and a Half-Cauchy prior for the beta-binomial concentration parameter κ. The model is fitted with Bambi (Capretto et al. 2022) on top of PyMC using the No-U-Turn Sampler (4 chains, 2,000 tuning iterations, and 1,000 draws per chain). Across all parameters, the maximum $\hat{R}$ was 1.020, and the minimum effective sample size was 320. Full implementation details are in the replication repository: https://github.com/WHaverals/quotable_canon.

[29] This distribution of compatible trajectories is the model's *posterior distribution*: the set of plausible values for each model parameter (b_i^0, b_i^1, $u_{a(i)}$, and so on), conditional on the observed data and the model's priors. Throughout the rest of this section, we summarize each poem's fitted prevalence in each decade by its posterior mean and a 95 percent credible interval. We describe a poem as "credibly rising" or "credibly declining" when the 95 percent credible interval for its slope b_i^1 lies entirely above or below zero, respectively.

[30] The controversy ran from 1819 to 1826 and turned on a distinction William Lisle Bowles had drawn in his 1806 edition of Pope: the suggestion that poetry which takes natural objects for its subject is intrinsically more elevated than poetry which takes artificial ones, of the kind Pope had so often worked with. Its ostensible subject was Pope's personal poetic standing; what it staged was a more fundamental disagreement over whether the heroic-couplet didactic mode could still be counted as poetry in the post-Romantic sense.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

Table 2. Top 15 credibly declining poems in the sustained-presence population, ranked by posterior-mean slope b_i^1 (log-odds per decade) against the corpus-wide trend

Rank	Work title	Author/Translator	b_i^1	95% CrI	LW
1	*Pastoral V* (transl. 1697)	Virgil/John Dryden	−0.27	[−0.33, −0.20]	31
2	*The Art of Poetry* (transl. 1680)	Horace/Wentworth Dillon, Earl of Roscommon	−0.25	[−0.30, −0.21]	92
3	*Prince Arthur* (1695)	Sir Richard Blackmore	−0.25	[−0.30, −0.19]	58
4	*Upon the Earl of Roscommon's Translation of Horace* (1680)	Edmund Waller	−0.24	[−0.31, −0.18]	32
5	*The Art of Poetry* (transl. 1757)	Horace/William Duncombe	−0.24	[−0.30, −0.18]	42
6	*The Georgics of Virgil* (transl. 1753)	Virgil/Joseph Warton	−0.24	[−0.29, −0.18]	50
7	*Medulla Poetarum Romanorum* (transl. 1737)	Henry Baker	−0.23	[−0.27, −0.20]	214
8	*Pastoral I* (transl. 1697)	Virgil/John Dryden	−0.23	[−0.29, −0.17]	45
9	*The First Satire of Persius* (transl. 1693)	Persius/John Dryden	−0.22	[−0.28, −0.17]	39
10	*An Essay on Translated Verse* (1684)	Wentworth Dillon, Earl of Roscommon	−0.22	[−0.26, −0.18]	125
11	*Pastoral III* (transl. 1697)	Virgil/John Dryden	−0.22	[−0.28, −0.16]	35
12	*Metamorphoses, Book XII* (transl. 1700)	Ovid/John Dryden	−0.22	[−0.28, −0.15]	31
13	*Metamorphoses, Book I* (transl. 1693)	Ovid/John Dryden	−0.21	[−0.26, −0.16]	48
14	*An Essay upon Poetry* (1682)	John Sheffield, Duke of Buckingham	−0.21	[−0.25, −0.16]	79
15	*Virgil's Æneid* (transl. 1740)	Virgil/Christopher Pitt	−0.20	[−0.23, −0.16]	140

Note: Credibly declining poems are those whose 95 percent credible interval on b_i^1 excludes zero on the negative side. *LW*, lifetime PPA host works.

Table 3. Top 15 credibly rising poems in the sustained-presence population, ranked by posterior-mean slope b_i^1 (log-odds per decade) against the corpus-wide trend

Rank	Work title	Author	b_i^1	95% CrI	LW
1	*Rime of the Ancyent Marinere* (1798)	Samuel Taylor Coleridge	+0.35	[0.27, 0.41]	287
2	*The Daffodils* (1807)	William Wordsworth	+0.31	[0.21, 0.40]	109
3	*Ode on Intimations of Immortality* (1807)	William Wordsworth	+0.30	[0.24, 0.36]	263
4	*Lines Written . . . Above Tintern Abbey* (1798)	William Wordsworth	+0.26	[0.19, 0.33]	170
5	*The Eve of St Agnes* (1820)	John Keats	+.24	[0.15, 0.33]	108
6	*Prometheus Unbound* (1820)	Percy Bysshe Shelley	+0.23	[0.15, 0.32]	128
7	*Is There for Honest Poverty* (1795)	Robert Burns	+0.21	[0.12, 0.30]	64
8	*Christabel* (1816)	Samuel Taylor Coleridge	+0.21	[0.14, 0.28]	157
9	*The Education of Nature* (1800)	William Wordsworth	+0.20	[0.11, 0.28]	74
10	*On His Blindness* (1673)	John Milton	+0.19	[0.12, 0.26]	99
11	*Peter Bell* (1819)	William Wordsworth	+0.19	[0.09, 0.28]	63
12	*To Daffadills* (1648)	Robert Herrick	+0.19	[0.10, 0.28]	55
13	*Sonnet 73* (1609)	William Shakespeare	+0.18	[0.11, 0.27]	70
14	*To a Waterfowl* (1818)	William Cullen Bryant	+0.18	[0.08, 0.28]	71
15	*Ae Fond Kiss* (1791)	Robert Burns	+0.18	[0.08, 0.27]	41

Note: Credibly rising poems are those whose 95 percent credible interval on b_i^1 excludes zero on the positive side. *LW*, lifetime PPA host works.

process in which older distinctions among verse kinds lost some of their force, and poetry came to be imagined more narrowly through the prestige of the lyric (Jackson 2005, 2016; LeGette 2017). The lower ranks show how that shift entered pedagogical use. Herrick's *To Daffadills* (1648), Shakespeare's *Sonnet 73* (1609), Bryant's *To a Waterfowl* (1818), and Burns's *Ae Fond Kiss* (1791) are short lyrics that anthologies could preserve, and pupils could learn by heart (Coiro 1990; Robson 2012).

Table 4 adds a third shape to this history. A small body of poems stayed put across the whole modeled window, holding their position through both the Augustan apparatus and the lyric-recitation regime that succeeded it.[31] What the model labels as stability is the durability of works that remained useful across changing regimes.

[31] Stability here is operationalized as a posterior-mean slope $|b_i^1| \leq 0.03$ together with a tight credible interval (width ≤ 0.15).

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

Table 4. Top 15 stable poems in the sustained-presence population, ranked by lifetime PPA host works

Rank	Work title	Author/Work	b_i^1	95% CrI	LW
1	*Psalms* (1611)	King James Bible	−0.011	[−0.03, 0.01]	1,110
2	*Proverbs* (1611)	King James Bible	+0.007	[−0.01, 0.03]	738
3	*Job* (1611)	King James Bible	−0.005	[−0.03, 0.01]	683
4	*Othello* (1623)	William Shakespeare	+0.017	[−0.01, 0.04]	661
5	*King Lear* (1623)	William Shakespeare	+0.021	[0.00, 0.04]	579
6	*The Faerie Qveene* (1596)	Edmund Spenser	+0.023	[0.00, 0.05]	560
7	*Henry VIII* (1623)	William Shakespeare	+0.026	[0.00, 0.05]	544
8	*Night Thoughts* (1742–46)	Edward Young	−0.003	[−0.03, 0.02]	526
9	*Henry IV Part 1* (1598)	William Shakespeare	+0.027	[0.00, 0.05]	499
10	*Essay on Man, Epistle IV* (1734)	Alexander Pope	−0.018	[−0.04, 0.01]	490
11	*King John* (1623)	William Shakespeare	+0.026	[0.00, 0.05]	467
12	*Essay on Man, Epistle I* (1733)	Alexander Pope	−0.019	[−0.04, 0.01]	462
13	*Henry IV Part 2* (1600)	William Shakespeare	+0.009	[−0.02, 0.04]	459
14	*Henry V* (1623)	William Shakespeare	+0.007	[−0.02, 0.03]	450
15	*Measure for Measure* (1623)	William Shakespeare	+0.020	[−0.01, 0.05]	379

Note: Stable poems are those whose posterior-mean slope $|b_i^1| \leq 0.03$ and whose 95 percent credible interval is tight (width ≤ 0.15) – poems pedagogical writers reached for repeatedly across the whole modeled window, neither gaining nor losing ground relative to the corpus trend. *LW*, lifetime PPA host works.

The King James Bible (1611) occupies the first positions in the table, with *Psalms*, *Proverbs*, and *Job* all remaining steadily available across the modeled window. A verse from *Proverbs* shows how that stability worked in practice:

> Train up a child in the way he should go:
> and when he is old, he will not depart from it.

Thomas Dilworth's *New Guide to the English Tongue* (1751) uses the lines to teach spelling and reading. Robert Ross's *American Grammar* (1782) and John Walker's *Rhetorical Grammar* (1785) cite the same verse, and more than a century later, John Baker Opdycke's *Composition Planning* (1913) identifies it as an example of parallel construction.

Shakespeare supplies the second great reservoir. A passage from *Othello* shows how dramatic speech became schoolroom material. Bysshe's *Art of English Poetry* (1702) already breaks the play into keyed examples. Several handbooks return to Iago's lines on reputation:

> Good name in man and woman
> Is the immediate jewel of their souls.

Pope's *Essay on Man* remains stable where it yields lines that can be lifted from the larger argument. A couplet from Epistle IV appears in Daniel Fenning's *New Grammar of the English Language* (1771) and Sheridan's *Lectures on the Art of Reading* (1775) and is still present in PPA works more than a century later. This mostly happens through couplets like these that do not require the rest of the poem around them:

> Honour and shame from no condition rise;
> Act well your part, there all the honour lies.

Figure 6 shows the fitted trajectories of selected poems. Each panel shows how a poem's quotation prevalence changed relative to the average modeled poem. The dashed line marks that poem's mean relative position over the decades shown; values above or below it indicate a stronger or weaker relative presence. This centers large and small presences on the same scale, so that *Paradise Lost* – the single most-cited work in the corpus by raw host-work count (Table 1) – can be read alongside poems entering prosody writing much later and more modestly. Dryden's *Pastoral V* and Coleridge's *Rime* are already familiar from Tables 2 and 3: the first traces the sharpest decline in the panel, and the second the sharpest rise. The more telling case here is James Thomson's *Winter* (1726). *The Seasons*, of which *Winter* was the first-published book, was one of the most widely read long poems of the eighteenth and early nineteenth centuries (St. Clair 2004), and through most of the modeled window, its first book sits close to its own mean – neither gaining nor losing ground – before bending into a gradual late decline. *Paradise Lost* peaks around the 1760s and recedes from there, even as the poem remains the most-cited text in the archive by raw host-work count (Table 1). The panel shows relative displacement: a poem crowded out at the margin by a broadening quotation economy while retaining absolute dominance. That trajectory sits inside a long reception history in which Milton's epic had already been made canonical through eighteenth-century editorial labor and commentary before Romantic and later critical discourse recast its cultural function (Kolbrener 2014). *Elegy Written in a Country Churchyard* (1751) and *The Deserted Village* (1770) trace the opposite movement, climbing slowly through the window from below their eventual means. Both were finding their footing after their initial publication moment. Goldsmith's poem was listed as a text for recitation in women's training colleges in the 1860 curriculum cycle, alongside *Paradise Lost* and Wordsworth's *The Excursion* (Robson 2012). That institutionalization came ninety

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

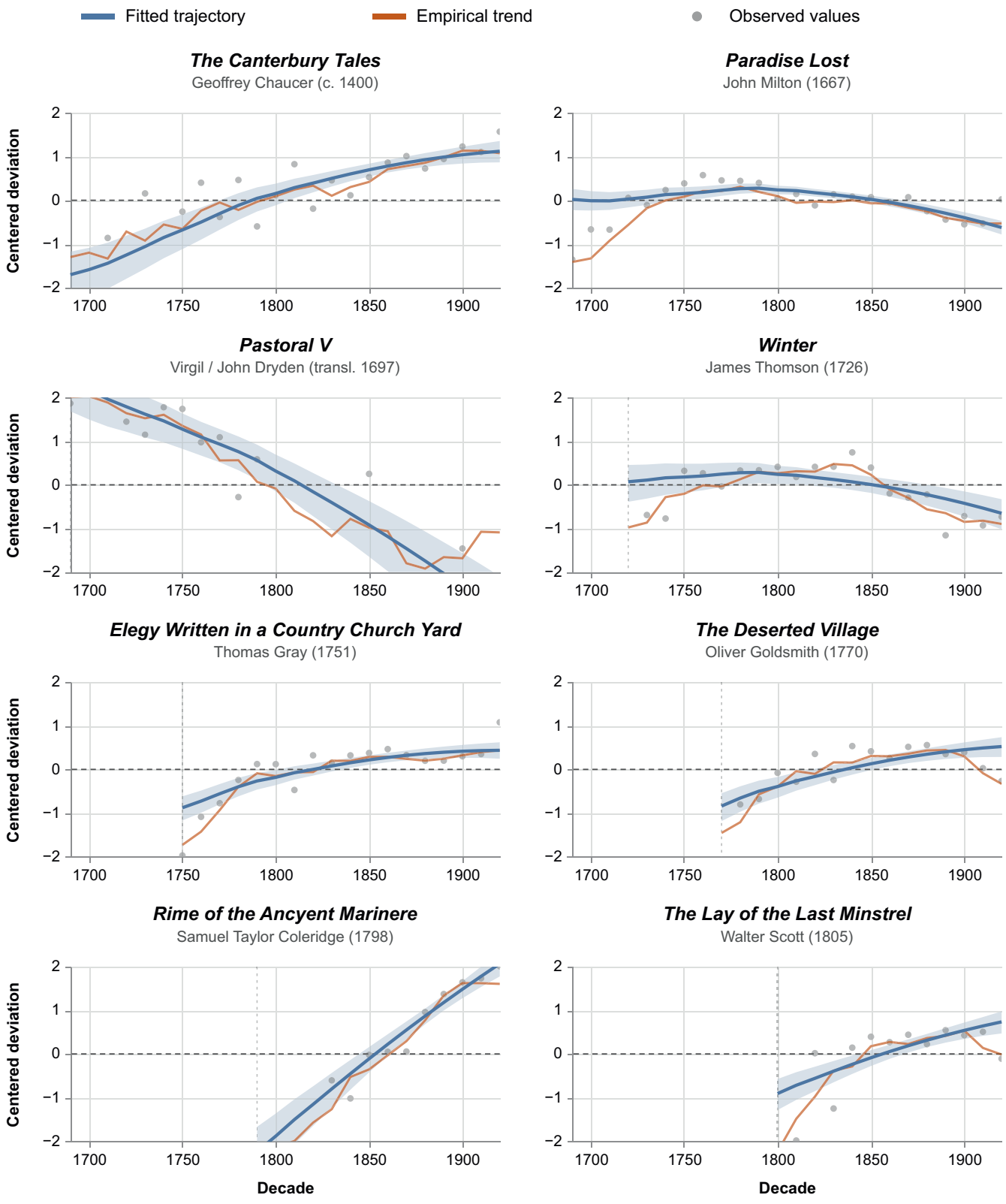


Figure 6. Selected poem trajectories, centered within poem. Each panel shows changes in a poem's quotation prevalence relative to the average modeled poem, on a log-odds scale. The horizontal dashed line marks the poem's mean fitted position over the decades shown: positive values indicate a stronger relative presence than this mean, and negative values a weaker one. The blue line shows the Bayesian fit; shading shows pointwise 95 percent credible intervals with the estimated baseline and poem mean held fixed. Gray points show observed deviations, and the orange line is a smoothed empirical comparison. Vertical dotted lines mark the publication decades. All panels use the same scales; portions of curves and credible bands, and observations, outside the vertical range are not shown.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

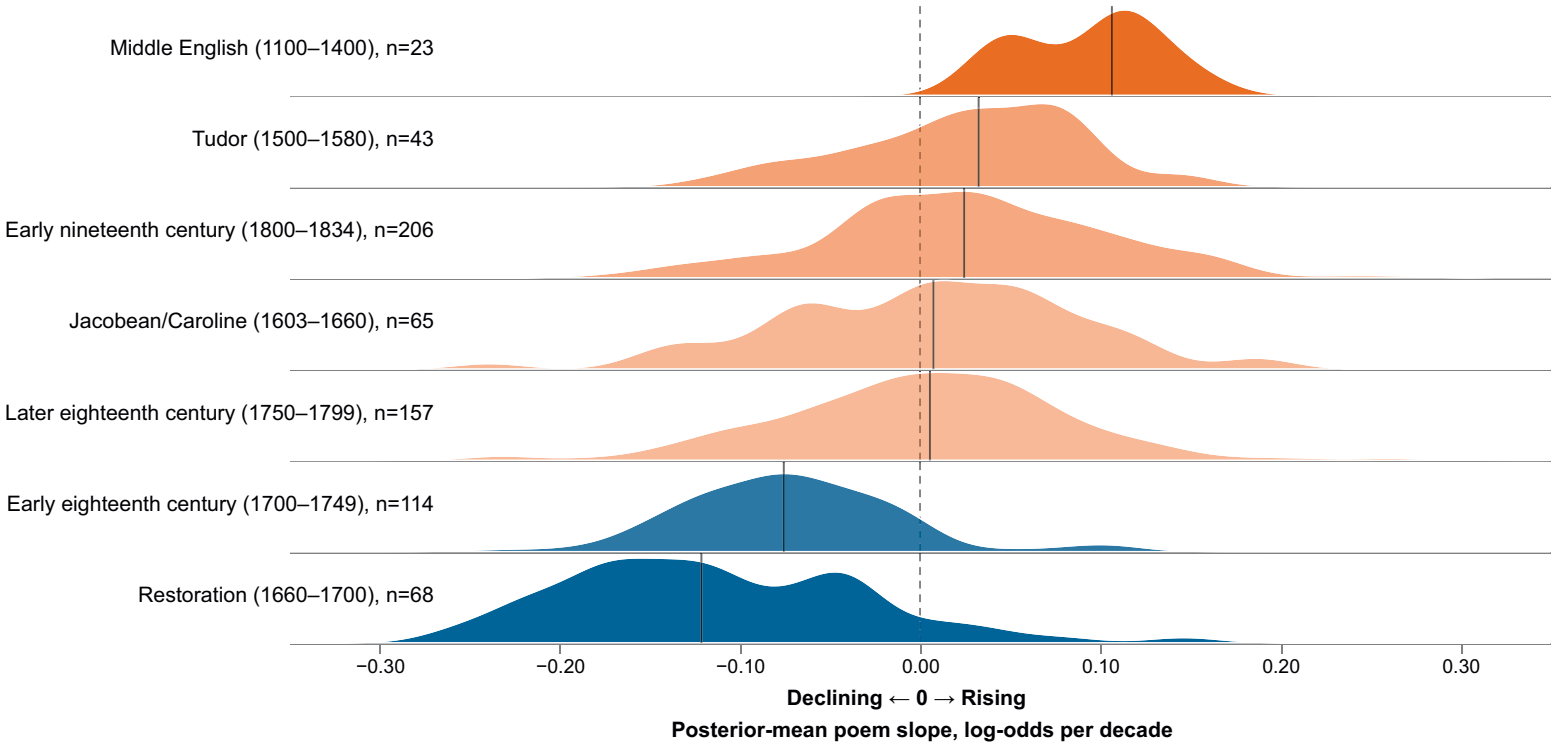


Figure 7. Distribution of poem-level slopes by Chadwyck-Healey literary period. Each ridge shows the distribution of posterior-mean poem slopes b_i^1 for poems assigned to that period, measured in log-odds per decade against the shared corpus trend. Values to the right of zero indicate poems gaining ground relative to the corpus-wide trend; values to the left indicate poems losing ground. The dashed vertical line marks zero, and the solid vertical line within each ridge marks that period's median slope. Rows are sorted by median slope and include only periods with at least 20 modeled poems.

years after its publication, and that timing is what the panel shows: a poem entering prosody writing well below its own eventual mean, then climbing as the curriculum caught up with it. *The Canterbury Tales* has the longest arc of all. Starting well below its own mean in the 1690s and rising steadily across the full two centuries of the modeled window, it traces the nineteenth-century construction of Chaucer as an object of literary history and national instruction (Utz 2020) – a poem that had to wait for a new kind of reader to find its place in prosody writing.

Periods rising and declining

The riser and decliner patterns visible at the poem level also show up when poems are grouped by the literary period to which they belong. Each poem in the population carries a period label from the Chadwyck-Healey metadata.[32] Figure 7 takes the same poem-level slope estimates that underlie Tables 2 and 3 and asks how those slopes are distributed within each period. Each ridge shows that distribution for one period.[33] The ridges are sorted from bottom to top by their median slope, so the figure can be read directionally: periods whose poems were collectively losing ground in the PPA sit at the bottom, and those whose poems were gaining ground sit at the top. The dashed vertical line marks zero – the corpus-wide trend – and the solid vertical line within each ridge marks that period's median.

The two bottom ridges – Restoration (1660–1700) and Early eighteenth century (1700–1749) – sit clearly to the left of zero: these are the periods that produced the top decliners in Table 2, and the figure shows a broad pattern of decline across the cohort. The Restoration ridge shows a pronounced internal split: two humps are visible. The left hump, centered around −0.15, is the classical-translation and didactic-verse apparatus: Roscommon's *The Art of Poetry* (−0.25), his *Essay on Translated Verse* (−0.22), Blackmore's *Prince Arthur* (−0.24), Dryden's translations of Persius (−0.22), Ovid's *Metamorphoses* (−0.21), Juvenal (−0.19), Sheffield's *Essay upon Poetry* (−0.20), and Garth's *Dispensary* (−0.18). Many of the poems named in Table 2 reappear here, and the cluster confirms that the decline was a collective fall of the entire apparatus. The right hump, centered near −0.05, is a different kind of Restoration poem: Dryden's original verse – *Absalom and Achitophel* (−0.05), *Mac-Flecknoe* (−0.04), and *Annus Mirabilis* (−0.03) – alongside Prior's *Henry and Emma* (−0.10) and Addison's *Campaign* (−0.15). These decline too, but more mildly, and most of their credible intervals straddle zero. What the bimodal ridge shows is that within the Restoration cohort, the translation apparatus fell hard, the original verse less so.

The Jacobean/Caroline ridge (1603–1660) also brings together poems with different trajectories. Its median sits just above zero, but the distribution extends to both sides of zero. The upper tail, rising clearly above zero, is almost entirely short lyric and dramatic song: Milton's sonnets, Shakespeare's individual sonnets and songs excerpted from the plays, Herrick's *To Daffadills* (+0.19), Lovelace's *To Lucasta* (+0.15), Jonson's songs to Celia, Herbert's *Virtue* (+0.12), and Marlowe's *Passionate Shepherd* (+0.09). The main body of the ridge, centered just above zero, is the Shakespeare plays and Milton's longer poems – *L'Allegro*, *Il Penseroso*, *Lycidas*, and *Samson Agonistes* – holding their position through the modeled window. Below zero, a declining tail: Dryden's adaptations of Jacobean drama, Denham's *Cooper's Hill* (−0.08), Cowley's *Davideis* (−0.14), the longer neoclassical narrative poems that belonged to the same tradition as the Restoration apparatus declining alongside them. The ridge is, in miniature, the lyricization process Jackson has described at the macro level (Jackson 2005, 2016): short lyric and song gained ground, plays and longer meditative poems held, and the heroic-narrative tradition gave way.

The Middle English ridge (1100–1400) lies furthest to the right. With only 23 modeled poems, the ridge is small, but its authors are the founding figures of the English literary tradition as the nineteenth century understood it: Chaucer above all, but also Langland's *Piers Plowman* (+0.12), Layamon's *Brut* (+0.13), Barbour's *Bruce* (+0.10), and Gower's *Confessio Amantis* (+0.06). Within Chaucer's own work, his dream-visions rose fastest: *The Book of the Duchess* (+0.16) and *The Legend of Good Women* (+0.15) sit a little above *Troilus and Criseyde* and the individual *Canterbury Tales* (which cluster together around +0.10).

Conclusion

The PPA holds four centuries of writing about verse, and scattered through it, in fragments, are the poems that writers reached for when they needed an example. This study begins to recover that record at scale: which poems were quoted, in which works, and what their trajectories through the writing looked like. That recovery was not achieved through simple tallying. A poem's presence in a host work is already an interpretation: of what counts as a quotation, of which poem is being quoted, of whether a passage that looks like Virgil is really Virgil or a later poet who has absorbed his lines. The host works are no less complicated. Anthologies copy from anthologies; a passage in circulation for fifty years may say less about contemporary taste than about the persistence of a "convenient" text. Rounce has called this the central problem of database-driven poetic reception study: a citation record is "a mechanism for producing data, rather than interpreting it," and "flattened out results lack context" (Rounce 2014). Simple counts would have flattened interpretive ambiguities into definitive-looking trends. We used statistical modeling to separate changes in quotation practice from the uneven growth of the archive and to estimate poem-level trajectories without treating sparse evidence as equally certain. Our entire pipeline – its filters, plausibility checks, and models – was a sustained attempt to interpret rather than merely produce.

That interpretive work begins with the corpora themselves. Our counts depend on two curated corpora: the PPA and the poetry corpus against which we read it. The PPA is not a neutral container for "prosody" as such, but a historically assembled archive of English-language works about language, poetry, and the places where the two overlap. In this archive, works of versification and elocution sit alongside grammar books, literary histories, school readers, anthologies, and miscellanies – genres that do not merely record a discourse on verse, but helped produce one. What we have traced, then, is a history of poetic quotation drawn from the works collected in the PPA and shaped by the standards of relevance that brought these works into relation with one another.

The Chadwyck-Healey poetry reference corpus introduces a second constraint. Because it is built around anglophone poetry, it lets us follow the English-language poems and translations that

[32] The labels are bibliographic categories that come with the corpus. They reflect the database's organizational logic rather than any particular literary-historical argument about periodization. One category, "Poems and Miscellanies, 1500–1900," is a catch-all for poems that could not be placed in a named period; we exclude it from our analysis here.

[33] We include only periods with at least 20 modeled poems; smaller cohorts are too sparse for a stable distribution.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

prosodic writers used, but it cannot tell the same story for every poetic tradition that entered the PPA. Latin and Greek citation, almost certainly heavy in early PPA works, is invisible to our pipeline; so are the French, German, and other modern-language poetic traditions that later PPA works occasionally draw on when making nationalist arguments about the distinctiveness of English versification. Both limitations are ones we can name but cannot remedy within the present study.

Within these limits, the evidence reveals a dense culture of poetic contact. *Passim* was useful precisely because it did not begin from titles, authors, or declared sources; it searched for shared wording wherever that wording appeared. But that strength also means that the alignments have to be read as evidence of contact before they can be read as evidence of quotation. The *King James Bible*, for example, emerges in our results as one of the PPA's major sources of poetic citation, but biblical language also travels through later poems, paraphrases, hymns, and commonplace phrases. The same is true of Shakespeare, Virgil, Gray, and other heavily sedimented sources. A match may indicate that a prosody manual quotes a canonical poem; it may also indicate that a minor poem in the reference corpus contains a canonical line, and that the PPA host is really reaching for the older source through the later one. What is historically the familiar condition of poetic intertextuality thus becomes, in our technical operationalization, a problem of attributional bleed-through or poetic "contamination." We can see this clearly in the example of Alaric Alexander Watts's elegy *To the Memory of George Barret* (1842). In the modeled data, the poem appears as a strong riser throughout the PPA; yet the signal is almost entirely carried by the canonical borrowings woven through it. Watts echoes Shakespeare's *Tempest* – "cloud-capped towers" and "gorgeous palaces" – and closes on Gray's *Elegy*: "leaves the world to darkness and to me." The quotation marks are Watts's own. He knows he is borrowing. *Passim* catches the shared wording too, but it cannot know, without interpretation, whether the historical object is Watts, Gray, Shakespeare, or the culture that made such lines available for reuse. Similar cases run through the long tail of the corpus: cento poems composed from borrowed lines, compilers such as Nathaniel Crouch reproducing other poets' verse under a pseudonym (Axon 1885). Together they remind us that, for much of the period studied here, the boundary between *quoting* a poem and writing *with* poetry was permeable.

Samuel Johnson's *Lives of the Poets* (1779–1781) is perhaps the most celebrated record of poetic authority being claimed and assigned. The PPA preserves the longer, much quieter record of that authority being used – or not. By the time a poet's lines were reaching prosody manuals, grammar books, and elocution treatises, something had already happened to them: they had entered the daily working life of a literary culture that needed examples. Recovering that record at scale, for the full range of poets whose lines circulated through four centuries of writing about verse, is what this study has attempted. The verdicts in the *Lives* shaped which poets entered that circulation. The record traced in our contribution follows what happened after poetic authority entered use: how poems became examples, and how those examples moved through the prose that taught readers how verse worked.

Data availability statement. Replication data and code are openly available. The Found Poems Dataset is archived at https://doi.org/10.5281/zenodo.20044417. The cleaned and reconciled poet metadata is archived on Zenodo at https://doi.org/10.5281/zenodo.20597001. The analysis code and computational notebooks are available at https://github.com/WHaverals/quotable_canon. The PPA full-text dataset on which the alignments were run is available at https://doi.org/10.34770/bze8-xk08 (Martin, Koeser, and Naydan 2026) and described in Koeser et al. (forthcoming).

Acknowledgements. We are grateful to Mary Naydan, who kept this many-handed effort moving, curated the supplementary reference corpus, and adjudicated the manual annotations used to evaluate *passim*. We also thank Anthony Nathan, Olivia Roslansky, and Molly Taylor for annotating poetry quotations in the evaluation set.

Author contributions. Conceptualization: W.H., M.M., R.S.K., and L.T.; Data curation: W.H., R.S.K., and L.T.; Formal analysis: W.H.; Methodology: W.H., R.S.K., and L.T.; Resources: M.M., R.S.K., and L.T.; Software: W.H., R.S.K., and L.T.; Supervision: M.M.; Visualization: W.H.; Writing – original draft: W.H.; Writing – review and editing: W.H., M.M., R.S.K., and L.T. All authors approved the final submitted draft. CRediT taxonomy: https://credit.niso.org/.

Funding statement. This research was supported by a Collaborative Research Grant from the Center for Digital Humanities at Princeton University and a grant from Princeton University's University Committee for Research in the Humanities and Social Sciences.

Competing interests. The authors declare no competing interests.

Use of AI tools. In preparing this manuscript, we made minimal, non-generative use of *Claude Sonnet 4.6* for spelling and language checking of author-written prose, review of author-written analysis code, and feedback on figure presentation. The research design, analysis, and interpretation are the authors' own. Every line of code has been checked, and the authors take responsibility for the final manuscript, code, and figures.

References

Axon, William E. A. 1885. "Burton's Books." *Book-Lore: A Magazine Devoted to Old Time Literature* 1: 129–37.

Benedict, Barbara M. 2015. "Choice Reading: Anthologies, Reading Practices and the Canon, 1680–1800." *The Yearbook of English Studies* 45: 35.

Blair, Hugh. 1783. *Lectures on Rhetoric and Belles Lettres*. Dublin: Messrs. Whitestone, Colles, Burnet, Moncrieffe, Gilbert [and 9 others in Dublin].

Bonnell, Thomas Frank. 2008. *The Most Disreputable Trade: Publishing the Classics of English Poetry 1765-1810*. Oxford: Oxford University Press.

Breen, Henry Hegart. 1857. *Modern English Literature*. London: Longman, Brown, Green & Longmans.

Brogan, Terry V. F. 1981. *English Versification, 1570-1980: A Reference Guide with a Global Appendix*. Baltimore: Johns Hopkins University Press.

Brooke, Stopford Augustus. 1880. *Theology in the English Poets*. London: C.K. Paul & Co.

Büchler, Marco. 2013. "Informationstechnische Aspekte Des Historical Text Re-Use." Ph.D. thesis, Universität Leipzig.

Büchler, Marco, Philip R. Burns, Martin Müller, Emily Franzini, and Greta Franzini. 2014. "Towards a Historical Text Re-Use Detection." In *Text Mining*, Theory and Applications of Natural Language Processing, edited by **Chris Biemann and Alexander Mehler**, 221–38. Cham: Springer International Publishing.

Capretto, Tomás, Camen Piho, Ravin Kumar, and Jacob Westfall. 2022. "Bambi: A Simple Interface for Fitting Bayesian Linear Models in Python." *Journal of Statistical Software* 103, no. 15: 1–29. https://doi.org/10.18637/jss.v103.i15.

Chadwyck-Healey, Charles. 2020. *Publishing for Libraries: At the Dawn of the Digital Age*. London: Bloomsbury.

Chandler, James. 1984. "The Pope Controversy: Romantic Poetics and the English Canon." *Critical Inquiry* 10, no. 3: 481–509.

Clark, John Scott. 1886. *A Practical Rhetoric for Instruction in English Composition and Revision in Colleges and Intermediate Schools*. New York: H. Holt and Company.

Coffee, Neil, Jean-Pierre Koenig, Shakthi Poornima, Roelant Ossewaarde, Christopher Forstall, and Sarah Jacobson. 2012. "Intertextuality in the Digital Age." *Transactions of the American Philological Association* 142, no. 2: 383–422.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.

Coiro, Ann Baynes. 1990. "Introduction: Robert Herrick and the Hesperides: On the Edge of the Renaissance." *George Herbert Journal* 14, no. 1: i–vi.

Compagnon, Antoine. 1979. *La Seconde Main: Ou, Le Travail de la Citation*. Paris: Seuil.

Cordell, Ryan. 2015. "Reprinting, Circulation, and the Network Author in Antebellum Newspapers." *American Literary History* 27, no. 3: 417–45.

Delpeuch, Antonin, Tom Morris, David Huynh, Stefano Mazzocchi, Jacky, Thad Guidry, elebitzero, Owen Stephens, Isao Matsunami, Albin Larsson, Iain Sproat, Silvério Santos, Allana Mayer, Kushthedude, Martin Magdinier, Luca Martinelli [Sannita], Sandra Fauconnier, Ekta Mishra, Antoine Beaubien, Lu Liu, Florian Giroud, Joanne Ong, Fabio Tacchelli, Allan Nordhøy, Elroy Kanye, Yaron Shahrabani, and Mathieu Saby. 2026. "OpenRefine/OpenRefine: 3.10.1." Zenodo. https://doi.org/10.5281/zenodo.18868465.

Gardiner, William. 1856. *The Music of Nature*. Boston: Sanborn, Carter & Bazin.

Gini, Corrado. 1912. *Variabilità e mutabilità: Contributo Allo Studio Delle Distribuzioni e Relazioni Statistiche*. Bologna: Cuppini.

Gladstone, Clovis, Russ Horton, and Mark Olsen. 2008. "TextPAIR: Pairwise Alignment for Intertextual Relations." ARTFL Project, University of Chicago. https://code.google.com/archive/p/text-pair/.

Guillory, John. 1993. *Cultural Capital: The Problem of Literary Canon Formation*. Chicago: University of Chicago Press.

Haverals, Wouter, Artjoms Šeļa and Petr Plecháč. forthcoming. "Canon Dynamics in the Princeton Prosody Archive, 1559–1929." *Journal of Cultural Analytics*. https://doi.org/10.22148/jca.1429.

Heuser, Ryan. 2025. "Generative Aesthetics: On Formal Stuckness in AI Verse." *Journal of Cultural Analytics* 10, no. 3: article 1036. https://doi.org/10.22148/001c.144825.

Hostetler, Selena, Mary Naydan, and Rebecca Sutton Koeser. 2023. "Streamlining Search Results with Clusters." Princeton Prosody Archive. https://prosody.princeton.edu/editorial/2023/04/streamlining-search-results-with-clusters/.

Hyde, Mary F. 1906. *Two-Book Course in English*. Boston: D.C. Heath & Co.

Jackson, Virginia. 2005. *Dickinson's Misery: A Theory of Lyric Reading*. Princeton, NJ: Princeton University Press.

Jackson, Virginia. 2016. "American Romanticism, Again." *Studies in Romanticism* 55, no. 3: 319–46.

Karlin, Daniel. 1995. "Introduction." In *English Poetry. A Bibliography of the English Poetry Full-Text Database*, iii–viii. Cambridge: Chadwyck-Healey.

Koeser, Rebecca Sutton. 2020. "Visualizing the Collections." Princeton Prosody Archive. https://doi.org/10.17613/vbf16-3d987.

Koeser, Rebecca Sutton, Mary Naydan, Laure Thompson, and Meredith Martin. forthcoming. "rinceton Prosody, from Archive to Dataset." Journal of Cultural Analytics. https://doi.org/10.22148/jca.1431.

Kolbrener, William. 2014. "Reception." In *The Cambridge Companion to Paradise Lost*, edited by **Louis Schwartz**, 195–210. Cambridge: Cambridge University Press.

Lane, Anthony. 1995. "Byte Verse. How to Wing Your Way Through Thirteen Hundred Years of English Poetry in an Afternoon of Interfacing." The New Yorker, 105–117.

LeGette, Casie. 2017. *Remaking Romanticism: The Radical Politics of the Excerpt*, Palgrave Studies in the Enlightenment, Romanticism and the Cultures of Print. Cham: Palgrave Macmillan.

Martin, Meredith. 2012. *The Rise and Fall of Meter: Poetry and English National Culture, 1860–1930*. Princeton, NJ: Princeton University Press.

Martin, Meredith. 2019. "Elizabeth Barrett Browning's Pathways through the PPA." Princeton Prosody Archive. https://doi.org/10.5281/zenodo.8292322.

Martin, Meredith. 2025. *Poetry's Data: Digital Humanities and the History of Prosody*. Princeton, NJ: Princeton University Press.

Martin, Meredith, Rebecca, Sutton Koeser, and Mary, Naydan. 2026. *Princeton Prosody Archive English-language versification and pronunciation (1532–1930): metadata, full text, and images*. Version 1. Data set. Princeton University. https://doi.org/10.34770/bze8-xk08.

Martin, Meredith, Meagan Wilson, Mary Naydan, and Rebecca Sutton Koeser. 2018. "Princeton Prosody Archive. Center for Digital Humanities at Princeton. Version 3.16.0." Accessed June 2, 2026. http://prosody.princeton.edu.

McCutcheon, Mark A. 2012. "The Cento, Romanticism, and Copyright." *ESC: English Studies in Canada* 38, no. 2: 71–101.

Mullen, Lincoln. 2015. "*textreuse: Detect Text Reuse and Document Similarity*." rOpenSci. https://github.com/ropensci/textreuse.

Mullen, Lincoln, and Yaoxiang Li. 2026. "*textreuse: Detect Text Reuse and Document Similarity*." https://docs.ropensci.org/textreuse/.

Nielsen, Mathias Wullum, and Jens Peter Andersen. 2021. "Global Citation Inequality is on the Rise." *Proceedings of the National Academy of Sciences* 118, no. 7: e2012208118.

Olsen, Mark, Russell Horton, and Glenn Roe. 2011. "Something Borrowed: Sequence Alignment and the Identification of Similar Passages in Large Text Collections." *Digital Studies/Le champ numérique* 2, no. 1. https://doi.org/10.16995/dscn.258.

Pocock, Stephen. 1996. "A Response to Dr Reimer's Review of the English Poetry Full-Text Database." *Computers & Texts* 11: 16–18.

Prins, Yopie. 2011. "'Break, Break, Break' into Song." In *Meter Matters: Verse Cultures of the Long Nineteenth Century*, edited by **Jason David Hall**, 105–34. Athens, OH: Ohio University Press.

Reimer, Stephen. 1996. "The Chadwyck-Healey English Poetry Full-Text Database." *Computers & Texts* 11: 13–15.

Robson, Catherine. 2012. *Heart Beats: Everyday Life and the Memorized Poem*. Princeton: Princeton University Press.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review* 71, no. 5: 845–58.

Rounce, Adam. 2014. "The Difficulties of Quantifying Taste: Blackmore and Poetic Reception in the Eighteenth Century." *Digital Defoe: Studies in Defoe and His Contemporaries* 6, no. 1: 19–35.

Salamone, Christopher. 2017. "'The Fragments, Scraps, the Bits and Greasy Relics': Shakespeare and the Eighteenth-Century Poetic Miscellany." *Eighteenth-Century Life* 41, no. 1: 7–31.

Scott, William. 1814. *Lessons in Elocution, or, a Selection of Pieces in Prose and Verse*. Boston: Isaiah Thomas, Jr.

Sheridan, Thomas. 1775. *Lectures on the Art of Reading*. London: J. Dodsley [and others].

Sheridan, Thomas, and John Henderson. 1796. *Sheridan's and Henderson's Practical Method of Reading and Reciting English Poetry*. London: E. Newbery.

Skeat, W. W. 1898. "On the Scansion of English Poetry." *Transactions of the Philological Society* 1895–1898: 484–503.

Smith, David A., Ryan Cordell, Elizabeth Maddock Dillon, Nick Stramp, and John Wilkerson. 2014. "Detecting and Modeling Local Text Reuse." In *IEEE/ACM Joint Conference on Digital Libraries*, 183–92. https://doi.org/10.1109/JCDL.2014.6970166.

St. Clair, William. 2004. *The Reading Nation in the Romantic Period* (Reprinted ed.). Cambridge: Cambridge University Press.

Thompson, Laure, Rebecca Sutton Koeser, Wouter Haverals, Mary Naydan, and Meredith Martin. 2026. "Princeton Prosody Archive Found Poems Dataset." Zenodo. https://doi.org/10.5281/zenodo.20044417.

Utz, Richard. 2020. "Chaucer among the Victorians." In *The Oxford Handbook of Victorian Medievalism*, edited by **Joanne Parker and Corinna Wagner**, 189–201. Oxford: Oxford University Press.

Vandenhoff, George. 1847. *The Art of Elocution*. New York: Shepard.

Downloaded from https://www.cambridge.org/core. 05 Nov 2026 at 09:27:28, subject to the Cambridge Core terms of use.