Enumerative ⓐ

Andrew Piper

Further Reading
Edited by Matthew Rubery and Leah Price

Abstract and Keywords

Enumerative reading is the attempt to account for quantity when we read. It entails a wide variety of practices that can range from accounting for the significance of numbers within texts, the number of texts themselves, to the numbers of words, or entities that populate texts. It activates three primary dispositions that each have significant genealogies within the humanities that I outline here. The first is self-reflexivity: enumeration makes us aware of the elements upon which our judgments and attachments to texts are made. Second, enumerative reading is translational in nature. It orients us to the movement between different sign systems. Finally, enumerative reading is synoptic. It aims to generalize and in so doing be active in the world. By way of illustration, I undertake an enumerative reading of the semantics of reading in a collection of ca. 6,500 novels published since the nineteenth century.

Keywords: reading, digital humanities, computational literary studies, cultural analytics, quantitative reasoning, intellectual history

Quantity has always been important to debates about reading. Since antiquity, advice manuals have warned against reading too much. Today, we worry about reading too little. Learning how to read the *right amount* is a venerable tradition in bibliotherapy. Numerology, or the study of the meaningfulness of quantity, has also been integral to textual interpretation since at least late antiquity. Augustine tells us that Peter caught 153 fish because it is the sum of 1 through 17, and seventeen equals ten (the number of commandments) plus seven (the number for the Holy Ghost). Lest one see this as a relic of a more superstitious age, Victor Hugo's epic of modern urban life, *Les Misérables*, contains 365 chapters, one for every day of the year, just as twelve-tone music provided the edifice upon which Thomas Mann would rewrite *Faust* to capture the horrors of the twentieth century. Quantity is a commonplace in books, not just in the content we read but also surrounding it: page numbers, chapter headings, volume numbers, publication dates, series numbers, call numbers, and ISBNs.

Enumerative

Enumerative reading is the attempt to account for quantity when we read, to bring together the sign systems of letters and numbers into a more integrated whole (as in *integer*, to make indivisible). It entails a wide variety of practices that can range from the significance of numbers within texts (why are there 1,001 nights?), to the sheer number of texts (what happens when Latin declines as a share of the European book market?), to the numbers of words or entities that populate texts (why do nineteenth-century novels share a predilection for a vocabulary of prevarication?). Sometimes the latter approach goes by the name distant reading, but this is a moniker that makes little sense in practice. Enumerative reading involves observing the most granular elements (periods, letters, words, phonemes, characters, syntax, numbers, sentences) as well as their most synoptic representation. It is deeply circular, or better yet, elliptical.

To enumerate means to count something out, to establish the number of something; in short, to *tally*. According to the *Oxford English Dictionary* (*OED*), *tally* derives from the Latin for rod or stick, often notched (in French we get *tailleur* or (p. 146) tailor). Tallying was used as a measure for taxation, or counting one's obligations. Quantity implies accountability—visibility and duration. We may love language, but quantity is more primordially related to inscription. Numbers have typically been more important to write down than words. Interestingly, the word tally is not thought to be etymologically related to the Germanic *tale*, which derives from the word for number (eventually becoming *Zahl* in modern German) and one of whose definitions is "to enumerate." As in numerous languages, there is a deep connection between the idea for counting and that of narration. When we read we are experiencing a recount.

Beginning in about the ninth century, it became common practice to tabulate the gospels by verse. "Canon tables," as they came to be known, provided a navigation tool to the New Testament, offering lists of which verses appeared in which gospels (all four, three, two, one). They were commonly illustrated surrounded by Byzantine archways. Quantity was the portal to theological understanding. By the twentieth century, counting the number of books in print would constitute the foundations of the new *Annales* school of social history. The number of readable things mattered.

One of the earliest ways that the quantity of words, rather than the quantity of books, was thought to be significant was in the work of Rudolf Flesch. Flesch was a Viennese immigrant who fled Austria from the Nazis and came to the US in 1933. He ended up as a student in Lyman Bryson's Readability Lab at Columbia University. The study of "readability" emerged as a full-fledged science in the 1930s, during the Great Depression, when the US government began to invest more heavily in adult education. Flesch's insight, which was based on numerous surveys and studies of adult readers, was simple. While there are many factors behind what makes a book or story comprehensible (i.e., "readable"), the two most powerful predictors are a combination of sentence length and word length. The longer a book's sentences and the longer its words, the more difficult readers will likely find it. Flesch reduced this insight into a single predictive, and somewhat bizarre, formula:

Enumerative

$$206.835 - 1.015 \times \frac{W}{St} - 84.6 \times \frac{Sy}{W}$$

W = number of words

St = number of sentences

Sy = number of syllables

According to Flesch's measure, Rudyard Kipling's *The Jungle Book* has a higher readability score (87.5) than James Joyce's *Ulysses* (81.0). Presidential inaugural speeches have been getting more readable over time. So too have novels. Ann Radcliffe's *Sicilian Romance* (63.5) is considerably more difficult than your average novel today. While there will always be exceptions—Hemingway's short sentences and pared-down vocabulary in *To Have and Have Not* is a good example of a text whose difficulty is not tied to its readability score—Flesch's measure is a remarkably successful way of assessing how difficult a piece of writing is. There are now well over thirty such measures, and the list continues to grow.

(p. 147) While Flesch's work went on to become highly influential within advertising and corporate communication, his greatest contribution lay in his best-selling book, *Why Johnny Can't Read*. It advocated the importance of phonics for developing readers. Flesch's book would become the inspiration for one of the most influential writers of the twentieth century: Dr. Seuss. For English speakers of my generation, quantitative knowledge lurks in the shadows of our most formative reading material.

Beginning in the 1980s with the rise of the field known as "natural language processing," and the subsequent dramatic increase in computing power in the 2000s, quantity would become increasingly important for our understanding of how we read. Premised on the single, straightforward idea of "distributional semantics," this work begins with the idea that language's meaning depends on the frequency of the contextual cues that surround it. Meaning, whether a single word, phrase, or an entire document, can be modeled, according to this view, as a probability distribution of co-occurring words (or sounds or any other kind of feature).

For example, my understanding of the word *bread* depends on the contexts in which I encounter it. Sometimes it will be framed as a food item, sometimes as a religious symbol, and from these "events" I develop a mental model of the word's various meanings. When *bread* is meant to convey a religious sense, it is more likely the case—though by no means guaranteed—that I will find some other theologically inflected language. Meaning is never deterministic. When I read the word *bread* in a new context, I activate this model and test it against the context that I am observing in that moment (just as I am also simultaneously updating my mental model to incorporate this new information). I do this at the document level, as well. When I read a work of fiction, I have a set of expectations about the signs I should encounter based on what I have already read. There can be deviations from my expectations, but overall I have a general sense of what I might expect to find. The more I read, the more sophisticated my model becomes and the more sensitive I am to the significance of variations. This is why in order to train people to be more analytical

when they read, we do so in part by asking them to read *more*. Quantity is an essential foundation to interpretation.

This process that I have just described is similar to how we now train machines to read. We enumerate, or tally, the "events" of features within a set of documents and then build a model that tries to approximate the category or ideas that we think they represent. It is important that, at least for now, we call this "reading," or perhaps *reading with*. Machines aren't reading by themselves, or achieving some kind of understanding of what they are reading. We are modeling the way that we think human minds model meaning. Enumerative reading is models all the way down.

Again, there is nothing new about this process. We have always used technology to help us read. The material structure of documents helps us know things in certain ways. The codex allows for random access (à la Augustine). The unadorned page that became more common around the turn of the nineteenth century allows for more immersive reading. The index allows for more non-sequential reading. (p. 148) And the birth of the critical edition allows for more genealogical reading, a better understanding of where a text comes from. Computational models are in this sense no different. They help us know things in certain ways (and not know other things). Despite their macrocosmic claims, computational models are perspectival.

So what are these perspectives? What are the *dispositions* of enumerative reading? I see three primary ones, though others may find more. The first is the way enumerative reading is highly self-reflexive. It turns my mind towards the units or elements from which my understanding of or attachment to a text is made. We might call this an awareness of the "lexical unconscious" after Walter Benjamin, who spoke of the optical unconscious with respect to photography.[1] Enumerative reading is very literal. When I engage in it, I explain how many books I have "read," how many features I have looked at, and how many things I have tested. There is a transparency to the machine's learnedness.

On the other hand, in many cases we cannot fully reconstruct how an algorithm arrives at the judgments it makes. As Michael Polanyi has argued, there remains a tacit dimension to knowledge, whether computational or bibliographic.[2] While we have spent a great deal of time worrying over the opacity of algorithms, we have spent far too little time reflecting on the obscurity of critical judgment. Enumerative reading gives us recourse to this self-reflexive knowledge of our judgments, even if necessarily incomplete. Enumerative reading allows me to try to know what I know.

Take for example the semantics of reading. One way to understand the history of reading is to model the semantic context of the word *read* in its various forms (*read*, *reads*, *reading*, *reader*, *readers*) and observe how these contexts have, or have not, changed over time.[3] What is it that people talk about when they talk about readers and reading? In the accompanying figure you can see a timeline of the words most commonly associated with these different forms of "read" when they appear in novels since the beginning of the nineteenth century down to the present (see Figure 12.1). The first thing you will notice is the remarkable stability of the semantics of reading. When people read in novels, not

much has changed. Those small changes, however, tell an interesting story. We can see an approximate moment towards the end of the nineteenth century when "letters" were no longer the most common thing that people read in novels. Books were. And they have stayed that way ever since, despite our imagined digitalness. The Bible too was also displaced, in this case by novels. Paper gains in significance as a medium, quite possibly as "the paper" (i.e., news). It has since been replaced in the present by "text." At the same time, "aloud" surprisingly remains one of the most common words associated with (p. 149) reading. Contrary to popular wisdom, silent reading is not the only or even the most common way reading is framed in fiction.
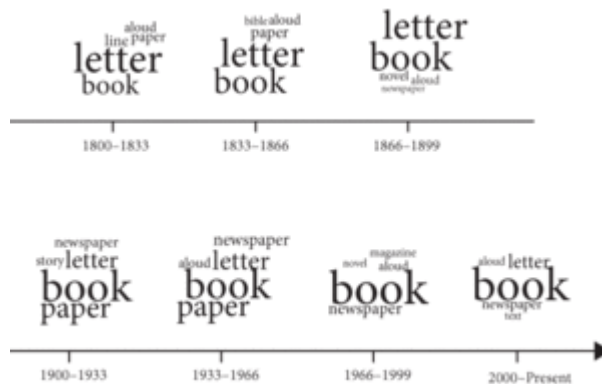

Figure 12.1 Most common words associated with forms of the word "read" in a collection of ~6,500 English-language novels.

Image Credit: Eve Kraicer-Melamed.

There are a host of other stories buried further down in these tables of words, which unfortunately I cannot show here but that are included in the supplementary data. For example, reading becomes considerably more religiously inflected during the mid-nineteenth century, while there is a new emphasis on "poetry" at the beginning of the twentieth. Today, in addition to "bedtime reading," a variety of manuscriptural objects, like diaries and journals, orbit around reading, suggesting the imaginary allure of handwriting in our electronic times.

These insights depend on how context is measured and, of course, which books and how many we take into consideration. Different data will tell a different story. But the process offers an opportunity to reflect on the infrastructures of our knowledge. Through quantity, we take into account the conditions of what we know and how we know it.

The second way enumerative reading disposes us towards our reading material is in its translational nature. Enumeration asks us to move between two different sign systems. In this, I see it as part of the long history of Humanism, which was in many ways founded upon the notion of linguistic and material difference. Humanism was about developing a set of reading practices that fostered the ability to understand the ways in which texts differ between times and places. Translation would emerge as one of its core practices as well as ideals. The knowledge gained in moving between languages, historical epochs,

and systems of writing was seen as the highest form of knowledge. Crossing the divide of textual and linguistic difference was (p. 150) a means of potentially crossing the divide to something more spiritually transcendent. Erasmus's bilingual New Testament, which was initially called a new "instrument," might be considered one of this tradition's most important founding documents.

Today, translating texts into quantities has emerged as the overwhelming feature of our cultural moment. Rather than see this as a kind of fallen state, I think we would do well to reposition it within this longer tradition of translational humanism, to see it as an ongoing attempt to understand the act of commensuration, of making different sign systems compatible with one another. Seen in this way, the literate and the numerate are not agons engaged in a duel, but two components of a more holistic understanding of human mentality.

Finally, enumerative reading disposes us towards our reading material in far more synoptic ways than other forms of reading. It is evident that when examining numerous texts we will move away from more particular observations. What has not been pointed out is that this happens even when we engage with just a few texts. Synoptic statements, not to mention titles, are woven into the fabric of critical discourse. They too are part of the infrastructure of our knowledge. According to Friedrich Schleiermacher, the so-called father of hermeneutics, these types of statements are not aberrant—illicit secretions on the discursive path towards knowledge of particularity—but essential indices of our understanding. As Schleiermacher consistently reminded us, knowledge of the part requires knowledge of the whole. We cannot *not* make such synoptic judgments. Even if we don't explicitly state them in our writing, they are informing our judgments. Enumerative reading foregrounds this process of generalization and subjects it to analysis. It models the conditions through which we arrive at synoptic statements about our reading material.

In his late novel, *Wilhelm Meister's Travels*, Goethe gives us an example of a new kind of modeler, the anatomical sculptor. Instead of being a dissector of dead things (called a *Prosektor* in Goethe's time), he envisions the idea of a *Proplastiker*, someone who makes in advance, who constructs for the future. In the paradox of the anatomical sculptor, Goethe combines the figure of the dissector, the analyst, with that of the builder, the synthesist. Rather than take individual bodies apart in a kind of second death, the modeler approximates analytical things to come. Analysis, the act of taking apart, and synthesis, the act of piecing together, are imagined to go hand in hand. According to Goethe, we cannot have one without a robust form of the other. This is enumerative reading's ellipse.

For Erasmus, one of Goethe's key predecessors, there was a crucially ethical dimension to synoptic reading. Paraphrase would emerge, perhaps surprisingly, as a vital practice in Erasmus' late output. It was designed, he said, to counter the intellectual partisanship of his age. As he wrote in the preface to his *Paraphrase on Acts*, comparing warring humanists to Europe's belligerent kings, "These chaotic enmities between one monarch and another, so fraught with disaster, so implacable, so long-continued, so far beyond all cure— are they not like some desperate sickness of the whole body? Can we discern any part of

the world that is immune from the infection of this dread disease? But even more destructive than that is this pestilence, (p. 151) which with its astounding and insoluble conflict of convictions has overmastered all men's minds."[4] The intellectual body politic for Erasmus was being taken apart by too much scholarly particularism. Paraphrase was to be the means through which the overly dissected text could be put back together, and in the process restore a sense of intellectual community. Beginning in 1517, Erasmus, who had already translated and edited the gospels, now turned to paraphrasing them, producing the longest running portion of his entire corpus.

Enumerative reading is the latest inheritor of this humanistic project. With its synoptic aims, it seeks to achieve a sense of community, a "common sense," to the act of interpretation. Through condensation it seeks consensus. Like the anatomical sculptor, it moves past the dissective literalism of the citation towards the synthetic ethos of the model. There is, and has always been, a distinct politics encoded in the act of enumeration.

Enumerative reading thus traverses two imaginary poles in the universe of reading. On the one hand, there is the Pythagorean dream of cosmic order that number signifies. We can find this, for example, in Dante's *Divine Comedy*, where, as Ernst Robert Curtius tells us, "100 cantos take us through three realms (33+33+33+1), the last of which contains 10 heavens. Triads and decads intertwine into unity."[5] When we think about and with numbers, order, form, and beauty are never far from view. On the other hand, there is also the diminutive *model* that number makes visible. We can see this, for example, on display in Goethe's "New Melusina," who lives in a traveling bookcase, or W. G. Sebald's Thomas Abrams, who builds a model of the Temple of Jerusalem for his whole life. Models are childish—playful, quixotic, but also a little bit subversive. Mischief is never far from play. Models tend to mock those who take their own pronouncements too seriously. They undo the certainty and rigidity of divine order.

But Sebald also wanted to remind us of the pain inherent in approximation as well as the potential joy of discovery that such pain entailed. The unreality of models can paralyze us if we get lost in them and never find our way out. But they can also open new worlds. "Now, as the edges of my field of vision darken," says Thomas Abrams, "I sometimes wonder if I will ever finish the Temple and whether all I have done so far has not been a wretched waste of time." After including an image of the Temple that covers both recto and verso and draws the reader's eye into the vanishing point of the book's spine, Sebald continues:

> But on other days, when the evening light streams in through this window and I allow myself to be taken in by the overall view, then I see for a moment the Temple...as if everything were already completed and as if I were gazing into eternity.[6]

Further Reading

Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary Study*. Ann Arbor: University of Michigan Press, 2018.

Enumerative

Hammond, Adam. *Literature in the Digital Age: An Introduction*. Cambridge: Cambridge University Press, 2016.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press, 2017.

Moretti, Franco. *Distant Reading*. London: Verso, 2013.

Piper, Andrew. *Enumerations: Data and Literary Study*. Chicago: University of Chicago Press, 2018.

Risam, Roopika, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston: Northwestern University Press, 2018.

Shore, Daniel. *Cyberformalism: Histories of Linguistic Forms in the Digital Archive*. Baltimore: Johns Hopkins University Press, 2018.

Turney, Peter D., and Patrick Pantel. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37 (2010): 141–188.

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press, 2019.

Notes:

[1] Walter Benjamin, "A Small History of Photography," in *One-Way Street and Other Writings*, trans. Edmund Jephcott and Kingsley Shorter (London: Verso, 1985), 225–39.

[2] "I shall reconsider human knowledge," writes Polanyi, "by starting from the fact that *we can know more than we can tell*." Michael Polanyi, *The Tacit Dimension* (Chicago: University of Chicago Press, 1966), 14.

[3] The code and data related to this project are located here: https://doi.org/10.7910/DVN/NGGAQ5.

[4] Desiderius Erasmus, *The Collected Works of Erasmus*, vols. 89 (Toronto: Toronto University Press, 1975), 50: 2–3.

[5] E. R. Curtius, "Numerical Composition," in *European Literature and the Latin Middle Ages* (Princeton: Princeton University Press, 1953), 501–509; 501.

[6] W. G. Sebald, *The Rings of Saturn*, trans. Michael Hulse (New York: New Directions, 1995), 238.

Andrew Piper

Enumerative

Andrew Piper is Professor and William Dawson Scholar in the Department of Languages, Literatures, and Cultures at McGill University. His work focuses on the relationship between the history of technology and reading. He is the author most recently of Enumerations: Data and Literary Study (2018).