



PROJECT MUSE®

---

## Debates in the Digital Humanities

Matthew K. Gold

Published by University of Minnesota Press

Matthew K. Gold.

*Debates in the Digital Humanities.*

Minneapolis: University of Minnesota Press, 2012.

*Project MUSE*. Web. 8 Feb. 2015<http://muse.jhu.edu/>.



➔ For additional information about this book

<http://muse.jhu.edu/books/9780816681440>

## Visualizing Millions of Words

MILLS KELLY

One of the very first posts I wrote for this blog was about visualizing information and some of the new online tools that had cropped up to make it a little easier to think about the relationships between data—words, people, and so on (Kelly). Interesting as they were, those tools were all very limited in their scope and application, especially when compared to Google’s newly rolled out Ngram viewer.<sup>1</sup> This new tool, brought to you by the good people at GoogleLabs, lets users compare the relationships between words or short phrases, across 5.2 million books (and apparently journals) in Google’s database of scanned works.

The data produced with this tool are not without criticism (Parry).<sup>2</sup> I will leave it to the literary scholars and the linguists to hash out the thornier issues here. My own concern is how using a tool such as this one can help students of the past make sense of the past in new or different ways. Among the many things I’ve learned from my students over the years is that they can be pretty persistent in their belief that words have been used in much the same way over time, that they have meant the same things (generally) over time, and that words or phrases that are common today were probably common in the past—assuming those words existed. They (my students) know that such assumptions are problematic for all the obvious reasons, but that doesn’t stop them from holding to these assumptions anyway.

I just spent an hour or so playing with the Ngram tool, putting in various words or phrases, and I can already imagine a simple assignment for students in a historical methods course. I would begin such an assignment by asking them to play with word pairs such as war/peace. By using Ngram, they would see that peace (red) overtook war (blue) in 1743 as a word that appeared in books in English (at least in books Google has scanned to date).

Intriguing as this “finding” is, the lesson that I would then focus on with my students is that what they are looking at in such a graph is nothing more or less than the frequency with which a word is used in a book (and only books) published over the centuries. While such frequencies do reflect something, it is not clear from one graph just what that something is. So instead of an answer, a graph like this one is a

doorway that leads to a room filled with questions, each of which must be answered by the historian before he or she knows something worth knowing.

After introducing my students to that room full of questions, I would then show them a slightly more sophisticated (emphasis on slightly) use of this tool. My current research is on the history of human trafficking. The term “human trafficking” (green) is a very recent formulation in books written in English. More common in prior decades were the terms “white slave trade” (blue) and “traffic in women and children” (red). This offers students a way to see the waxing and waning of these formulations over the past century.

But this also demonstrates a nice lesson in paying attention to what one is looking at. Google’s database of available books runs through 2008. The graph I describe ends in 2000. If I expand the lower axis to 2008, the lines look quite different. My hope would be to use tricks like this to demonstrate to my students how essential it is that they think critically about the data being represented to them in any graphical form.

While I doubt that I’ll ever assign Edward Tufte’s work to my undergraduates, I do think that an exercise such as this one with the Ngram viewer will make it possible to introduce the work of Tufte and others in a way that will be more accessible to undergraduates. If they’ve already played with tools like the Ngram viewer, then the more theoretical and technical discussions will make a lot more sense and will seem a lot more relevant. I think they will also be more likely to see the value in what Stephen Ramsay calls the “hermeneutics of screwing around.”<sup>3</sup>

## NOTES

This chapter originally appeared as “Visualizing Millions of Words” (<http://edwired.org/2010/12/17/visualizing-millions-of-words/>).

1. <http://ngrams.googlelabs.com/>.
2. <http://chronicle.com/article/Scholars-Elicit-a-Cultural/125731/>.
3. <http://library.brown.edu/cds/pages/705>.

## BIBLIOGRAPHY

- Kelly, Mills. “Visualizing Information.” *Edwired*. October 25, 2005. <http://edwired.org/2005/10/25/visualizing-information/>.
- Parry, Mark. “Scholars Elicit a ‘Cultural Genome’ From 5.2 Million Google-Digitized Books.” *Chronicle of Higher Education*, December 26, 2010. <http://chronicle.com/article/Scholars-Elicit-a-Cultural/125731/>.
- Ramsay, Stephen. “The Hermeneutics of Screwing Around.” Playing With History Conference. THEN/HiER. Niagra-on-the-Lake, Canada. April 17, 2010. <http://www.playingwithhistory.com/wp-content/uploads/2010/04/hermeneutics.pdf>.