

The N-Gram, The Witch, and the Wardrobe.

Contents

[Observing words through time.](#)

[Pluto Heading](#)

[Witch and Wizard Comparision](#)

Observing words through time.

An n -gram, not to be confused with a memory engram, the Ingram machine pistol or a scientific unit of mass, is defined as a contiguous sequence of some n items from text or speech, in other words, words. So now you should be asking yourself, great, but what can I do with that? Well think about the concept of a concordance for a book or an author's work. A concordance would be an element of the n -gram information for that word, but a real n -gram investigation would go well beyond just a book or just an author, and instead would be looking at everything that used that word. Sounds impossible, but it's not.

Google analytics is a tool that anyone can use to analyze word usage in books over time, at least books between 1800 and 2000. The Google Book's Ngram Viewer (<http://books.google.com/ngrams>) allows users to investigate a word's use and see the percentages and data for how often that word or phrase appears in print. The books being used in Google Books now included over 15 million scanned print edition and over 500 billion words. This means that we have a new element that can be included in a word study. It is also a way to look at culture over time by its word usage, and can be a way to look at how words change over time in their use.

Here is an example of a word being analyzed. As an old science teacher, I thought that I would start with "Pluto." Since recently the planet Pluto was designated from a major planet object to a



Milton- Odd Page Header

dwarf planetary body. When I go to the Ngram Viewer and type in the word “Pluto” as the search word and click on the Search button I get a graph of the word usage over time (see Figure 1). Now I can see how often that word appears in print over time for the books scanned. Notice the spike that occurs before 1940 - that would make sense as its discovery was in 1930’s. Also notice the rise in occurrences after 1980, when the debate was brewing about whether Pluto was a planet or not. I’m sure if the data continued, we would see another spike around 2006 when Pluto was reclassified. We can also see from the graph, that it isn’t a high use word either, as it maxed at about 0.00018% of word use.

Pluto Heading

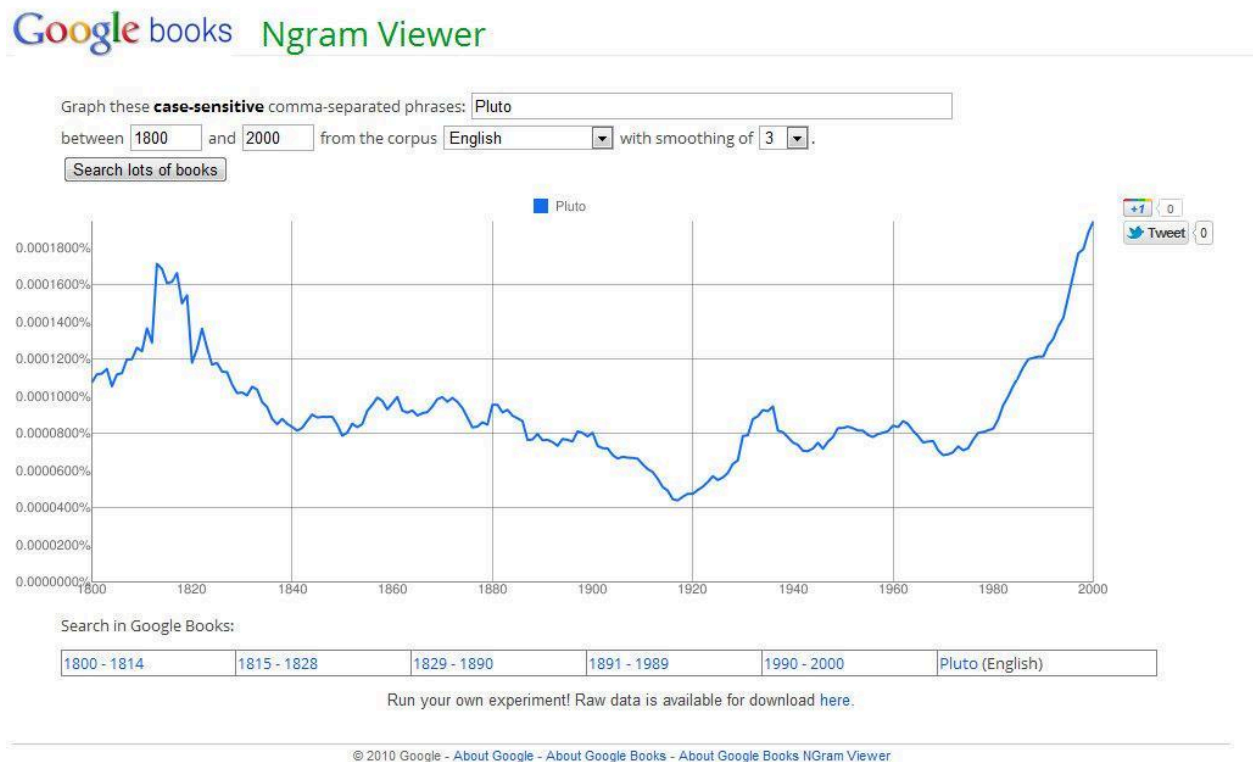


Figure 1: *n*-gram of the word “Pluto” from 1800 through 2000.





You can also use the system to compare words. Now instead of just typing in a single word into the search bar, you type in different words separated by a comma. Since the *Harry Potter* series just finished and Pottermore has yet to start let's consider some of the words used in that series. Upon researching I found that the word "wizard" occurs over 240 times in *Harry Potter and the Deathly Hallows* (although I'm sure that I could have missed a few in my search though the book). And thinking back as a child, I didn't hear much about wizards, but I often did hear about witches. So let's examine the word use of "witch" and "wizard" over time. Typing in "witch, wizard" into the search and searching gives me a graph display of the two words' appearances over time (see Figure 2). Looking at the chart we can easily see that indeed the word "witch" occurs in print much more often than "wizard," actually at least three times as much. But also in looking at the word wizard though time we might actually see the impact that J. K. Rowling had, as between 1980 and 2000, the word wizard starts becoming more common.

Witch and Wizard Comparision





Milton- Odd Page Header

Google books Ngram Viewer

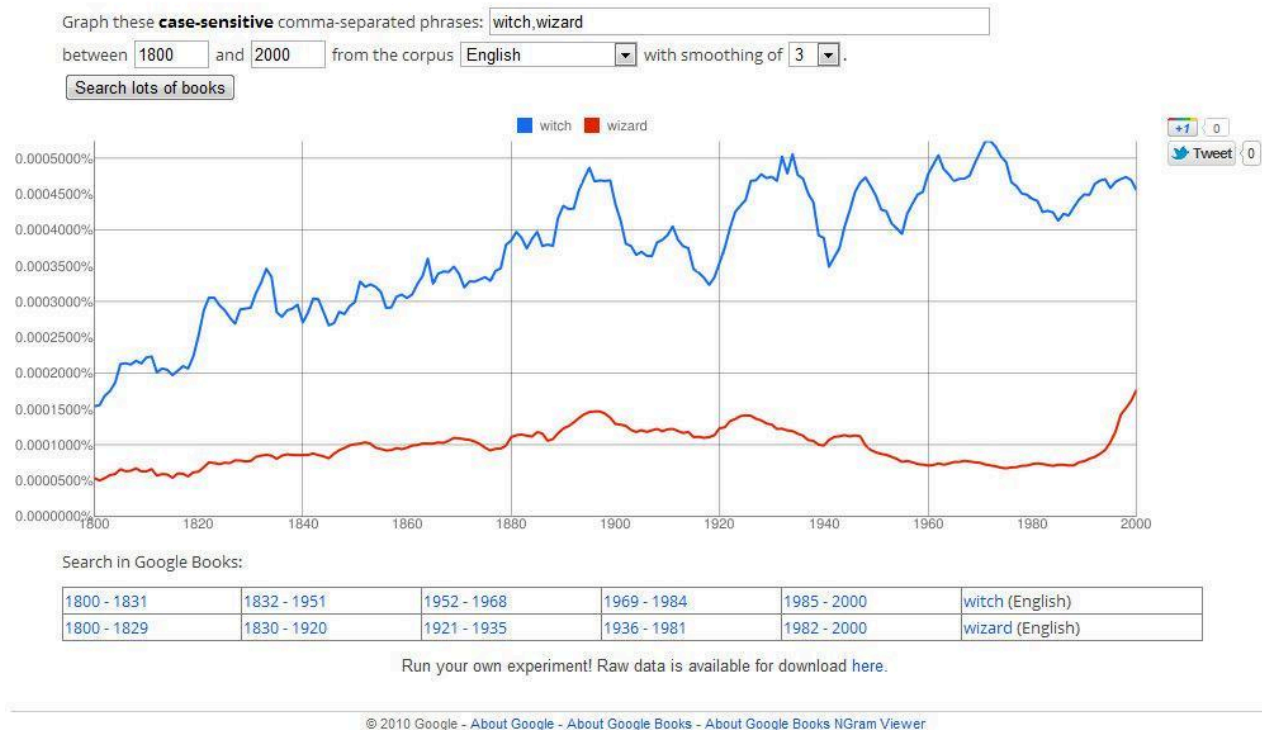


Figure 2: *n*-gram comparison of the words witch and wizard from 1800 through 2000.

One issue with the ngram viewer tool is that the words are case sensitive. Look at what happens when I do the same search as before, except capitalize the first letter (see Figure 3). Here while the word “Witch” is usually more common than “Wizard,” wizards really take off around the 1990s surpassing the word Witch by nearly double in 2000. Although when you examine the percentages you find that in 2000, the capital version of wizard was 0.0003% and the lower case was 0.00015%, and that witches were 0.00045% in lower case and 0.00016 capitalized, totaling out for wizards then at 0.00045% and witches with 0.00061% - so witches are indeed more common than wizards (at least so far).





Google books Ngram Viewer

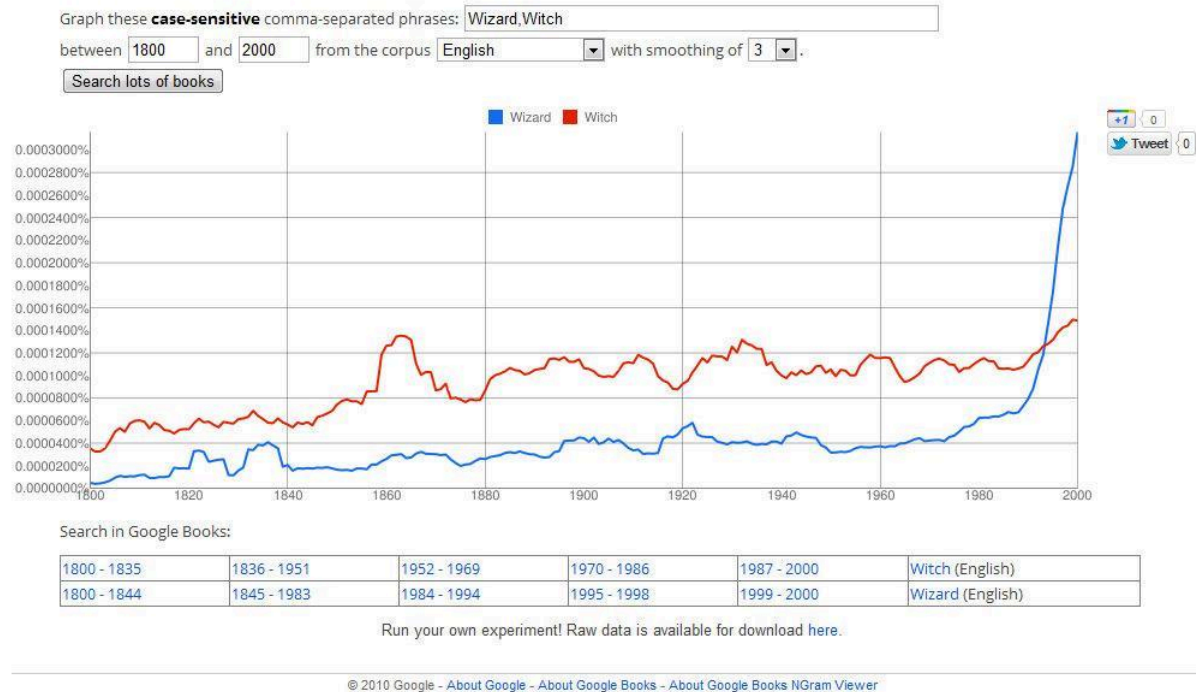


Figure 3: *n*-gram comparison of the words Wizard and Witch from 1800 through 2000.

I was also just recently reading that zombies are the new vampires, that vampires are going out and zombies are the next big thing - something that I'm starting to see in book publications. So I ran a comparison between those two terms (see Figure 4). Indeed it looks like zombies are becoming more common than they use to be and that around 2000, vampires shifted from a rising word to a declining use word.





Google books Ngram Viewer

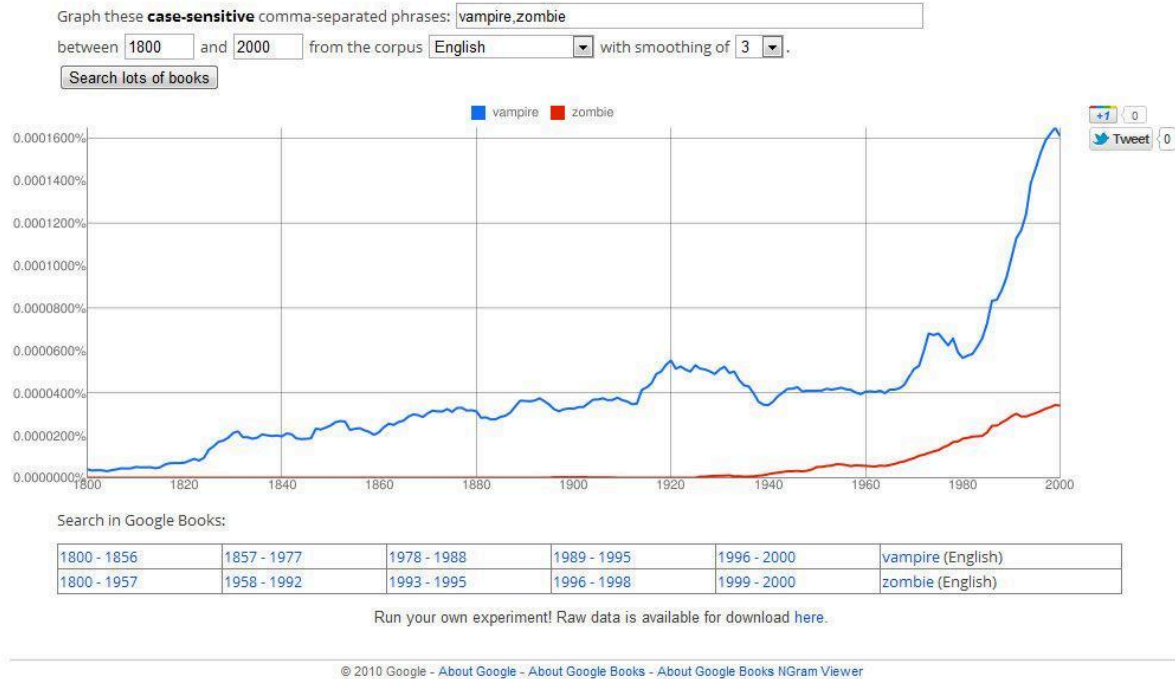


Figure 4: *n*-gram comparison of the words vampire and zombie from 1800 through 2000.

Recently a college student reading a short story based on the Cow Jumped Over the Moon, had a problem with the word “dryly.” She just didn’t know what the word meant. While she accepted and understood the definition when told, it made me wonder. So I researched the *n*-gram for the word “dryly” and got its graph (see Figure 5). That word was twice as common around 1900 as it is today (0.00006%). And while I wasn’t reading in 1900, it does help me to understand that she hasn’t been exposed to as many instances to a declining use word as I have.





Google books Ngram Viewer

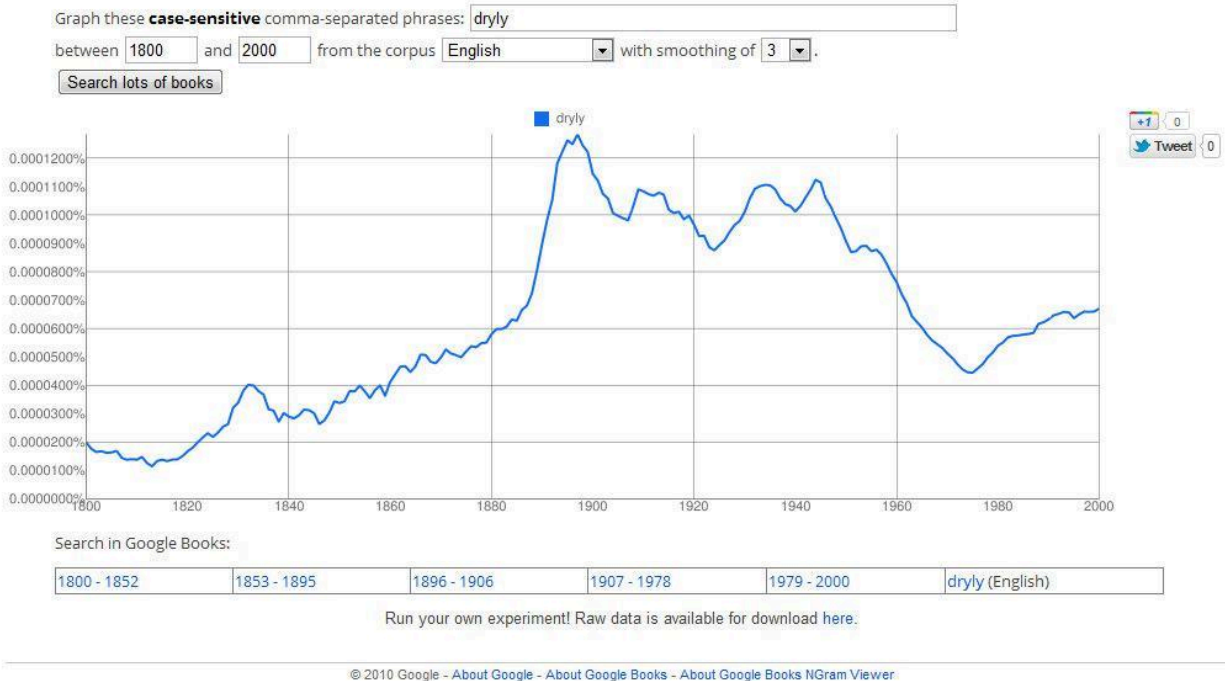


Figure 5: n-gram of the word “dryly” from 1800 through 2000.

The Ngram Viewer from Google, give educators a new tool in word research. Think about having your students search for words from your word wall or the books that they are reading and talk about how that word’s usage has changed though time (good math and science skill - reading graphs). You can also use the graph as a nice visual display about some words in the book that they are reading, such as for your bulletin board, letting them see the word use though time. This tool gives us an ability to do a different kind of Word Analysis, instead of phonics, decoding, or root words, we can instead look at how often people actually use the word in print.

Happy Reading with Technology.





Milton- Odd Page Header

