

“Average number of words
in the written vocabulary
of a 6- to 14-year-old
American child in 1945:

25,000

“Average number today:

10,000”

So reported *Harper's* in August 1990.¹ The validity of this comparison has been called into question as being one of not apples to apples, but rutabegas to rutabegas,² yet our curiosity about the matter remains: *How has our vocabulary changed over time?*

I propose comparing the aggregated vocabulary of the top 20 novels at Project Gutenberg³ with that of the top 20 books on the *New York Times* hardcover fiction best-seller list, and—in a manner informed by data visualizations such as those at Wordnik⁴ and Book Scraper⁵—to present those words that are shared between the two sets, as well as those words that are exclusive to one or the other. The project will show how the frequency of shared words has changed, as well as the breadth of each group's vocabulary overall.

Sample visualization from Book Scraper



Sample visualization from Wordnik of the word “vanity”



1 <http://is.gd/WaF>

2 <http://is.gd/WbZ>

3 <http://www.gutenberg.org/browse/scores/top>

4 <http://www.wordnik.com/>

5 <http://labs.timesonline.co.uk/bookscraper/>