

Exercise 1: Culturomics

Watch the YouTube Film by Michel & Lieberman, read the paper (below).

Then try the Google N-gram viewer (<https://books.google.com/ngrams/>) to see what you can learn by comparing the frequency of words over time. One example might be to compare how often *Deutschland* (and related terms) are mentioned as opposed to *Schweiz*, etc.) You can do this by searching for the following:

schweizer+schweizerisch+schweiz+schweizerische+schweizerisches,deutschland+deutscher+deutsche+deutsches

That is, you're comparing two terms separated by a comma, where each term may have several components joined by '+'.

Try some terms and comparisons. Here are some ideas (for American English).

1. One of these: slavery + slave + slaveholder; colonial + colonialism; liberal vs. neoliberal, incl. the terms with -ism; Populist + populism; Bohemian vs. hippie vs. beatnik, ...
2. negro vs. Afro-American vs. African American vs. black American vs. black (note that you'll need to look up how to restrict the search with part-of-speech tags to handle the noun 'black')
3. ... topic of your own choice (no need to check w. instructor on suitability, but please choose nothing that would be offensive in the discussion)

Exercise: Try some other examples of your own perhaps in your own language in order either (i) to show off how revealing such comparisons may be or (ii) to show the limits of such views.

Write up your evaluation of the exercise in 100-200 wd., and turn this in on Oct. 31 – presenting either (i) what it shows about how language use reflects the social issues; or (ii) how limited word (and phrase) frequency is as a gauge of cultural attention. Include a graph by printing the page to .pdf then taking a snapshot of the page. Be prepared to present your findings in 1-2 overhead sheets on **Nov. 5**

You may work on the exercise together, but please write up the evaluation by yourself.

Due Nov. 1st. Extra: volunteer to debate the thesis, either affirmative or negative:

Large-scale text analysis reflects the cultural attention of its time.

Literature

Michel, Jean-Baptiste & Erez Lieberman Aiden (2011) "What we learned from 5 million books" (Ted Talk) https://www.ted.com/talks/jean_baptiste_michel_erez_lieberman_aiden_what_we_learned_from_5_million_books

Michel, Jean-Baptiste, et al. (2011) "Quantitative analysis of culture using millions of digitized books." *Science* 331.6014: 176-182.