

Swarthmore College

Works

Linguistics Faculty Works

Linguistics

Fall 2016

Getting Started With COCA

K. David Harrison

Swarthmore College, dharris2@swarthmore.edu

Jamie A. Thomas

Swarthmore College, jthomas6@swarthmore.edu

Follow this and additional works at: <https://works.swarthmore.edu/fac-linguistics>

Let us know how access to these works benefits you

Recommended Citation

K. David Harrison and Jamie A. Thomas. (2016). "Getting Started With COCA". *Anthropological Linguistics/Linguistic Anthropology*. DOI: 10.24968/2476-2458.ling.208

<https://works.swarthmore.edu/fac-linguistics/208>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in Linguistics Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.

Due Thursday

10/4

Exercise: Getting started with COCA contemporary English

Part 1: Pre-corpus work

one sentence on why you decided?
Why it connects to social change?

- Decide what you will search
- Come up with a word or phrase that you are interested in, and say why; or think of a pair of terms that are similar in meaning and whose semantic spaces may reveal something about culture (more advanced)
 - Examples: *bossy* is a gendered term; *lady* vs. *woman*
- Form a hypothesis about your search term(s).
 - Example: My hypothesis is that *bossy* is a gendered term
- Write a short paragraph about what you expect to find and why.

Part 2: Proceed to the corpus


- Open a browser on a computer or tablet and go to the Corpus of Contemporary American English: <http://corpus.byu.edu/coca/>
 - If your question has to do with the history of a word or phrase, go to the Corpus of Historical American English: <http://corpus.byu.edu/coha/>
- Click on the Login icon on the upper right of the page and register (create username and password)
 - You are ready to go!

Part 3: Search the corpus

one phrase per person in group

We will first explore what you see with the default search setting, List View

- Keep all the settings at their default, and type in your word; click **Find matching strings** to search
 - This will bring you to the FREQUENCY tab
- How many times does the word occur in the corpus?
- To see what contexts the word occurs in, click on your search word shown in all caps under the **Context** button (what happens if you click the **Context** button? Do you know why?)

- This brings you to the CONTEXT tab
- Click the Help button, which is the question mark next to the login icon. What do you learn from this?
- What does this icon  stand for?
- Go back to the Context tab. For any individual token, how do you see the larger context that it occurs in?
- Go back to the Search tab
- Change the search view setting from **List** to **Chart**
- Search the corpus by clicking **See frequency by section**
- What genre is your word most frequent in? Do you see any trends?
- Consider the top row, SPOKEN. What are the numbers under **FREQ**, **SIZE (M)**, and **PER MIL**?
- What's the result if you divide **FREQ** by **SIZE (M)**?
- Next, look at the different time period bins. What time period is your word most frequent in? Is it increasing or decreasing in use?
- In general, your first few corpus searches will not yield the best results for your research question; keep refining your search! Keep track of the refinements that you make.

Part 4: Saving the data for offline analysis

- Once you have a search that you think brings up some useful data, you can save the data for offline analysis
- Open up a spreadsheet program like Excel and create a new document
- Go back to the Context view of your search
- Select and copy the results
- Paste the results into Excel (if you are given the option, paste as text)
- Clean up the spreadsheet as necessary
 - You might want to remove unnecessary rows
 - You might want to add rows for coding the data

what's your process? — how did you refine?

some sort of chart w/ frequency of tokens/types.