Corpus of Historical American English
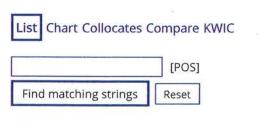
SEARCH   FREQUENCY   CONTEXT   OVERVIEW

List Chart Collocates Compare KWIC

[ ] [POS]

Find matching strings   Reset

☑ Sections Texts/Virtual Sort/Limit Options

In addition to this online interface, you can also download extensive data for offline use -- full-text data, word frequency, and n-grams data.

The Corpus of Historical American English (COHA) is the largest structured corpus of historical English. It is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English. If you are interested in historical corpora, you might also look at our Google Books (see comparison), Hansard, and TIME corpora.

COHA contains more than 400 million words of text from the 1810s-2000s (which makes it 50-100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade. The creation of the corpus results from a grant from the National Endowment for the Humanities (NEH) from 2008-2010.

Click on any of the links in the search form to the left for context-sensitive help, or take a look at the expanded overview of the corpus. You might pay special attention to the comparisons between decades and the (new) virtual corpora, which allow you to create personalized collections of texts related to a particular area of interest.

Five minute tour
Articles on using COHA for research
Compare to Google Books
Compare to small corpora (ARCHER, Brown family, etc)