ANALYZING TEXTS

You can entire texts (for example, texts from the Web, or a paper that you've written), and then see detailed information about the words in the texts, as well as selecting phrases in your text and then finding similar phrases in COCA. To access this feature, just click on the “text” icon at the top of the corpus (indicated in red below) and then paste in your text. In this overview, we are using a text about COVID from CNN.com.

Once you input the text, you will see the text with the words highlighted [1], with less frequent words (overall in COCA) in yellow. [2] shows the percentage of words from different frequency bands in COCA; the more in yellow, the “harder” the text may be. The most frequent words in each of these three frequency bands are shown in [3].
You can click on any **word** in either the text or the frequency/keyword lists (as with vaccine above) to see a detailed “word sketch” of that word (this is just a partial view of the page for vaccine).

**TOPICS** (more)
disease, virus, vaccination, infect, outbreak, flu, infection, epidemic, vaccinate, immunity, immune, infect, dose, infectious, injection, antibiotic, immunization, influenza, complication, transmit

**COLLOCATES** (more)
NOUN flu, influenza, dose, autism, virus, polio, disease, safety
VERB cause, receive, develop, prevent, protect, produce, contain, administer
ADJ effective, available, safe, pneumococcal, live, seasonal, oral, experimental
ADV currently, eg, genetically, routinely, safer, ie, commercially, ideally

**RELATED WORDS**
vaccinate

**SYNONYMS**
inoculation, preparation, serum, shot, vaccine

**CLUSTERS**
vaccine

**TEXTS / VIRTUAL CORPORA**

From this “home page”, you can access many **external resources**, such as definitions, etymology, images, videos, pronunciation, and translations.
This page contains links to more detailed pages based on corpus data, such as collocates, related topics, word clusters, texts to create Virtual Corpora, and concordance lines.
You can also select **phrases** (or parts of phrases) in your text, to find related phrases in COCA. This is like a “phraseological thesaurus”, since related patterns can affect how we perceive a given phrase. For example, we could select [1] the phrase **stunningly effective** or **note of optimism** or **steady drumbeat of**. The corpus inserts the words into the phrase sequence [2], and then we can modify this via the options in [3], such as finding all forms of a word, words with the same part of speech, synonyms, etc. We then click on [SUBMIT] to find related phrases in COCA.

No other corpus website allows users to do so much in terms of analyzing texts. English-Corpora.org is the only website that allows users to see an incredibly wide range of information on each word in the text, to help with reading comprehension. And as far as writing, no other website allows users to as easily and quickly look at related phrases to see patterns in the language.