

COCA (CORPUS OF CONTEMPORARY AMERICAN ENGLISH) BASICS

1. Getting started

Go to: <http://www.americancorpus.org>. You need to register (for free!) first.

2. Basic functions

A. Display Options

- **LIST**: Show a list of word(s) or combination of words (ranked according to their frequency)
- **CHART**: Show a chart comparing frequencies of a word in different genre or time.
- **KWIC**: Show the key word(s), i.e. search word(s), in contexts
- **COMPARE**: Compare two words according to their frequencies (just generally or with a certain collocate)

B. Types of queries (Search string)

- **WORD**: a search word or phrase
- **COLLOCATES**: a word (not a phrase) that occurs within up to 10 words before / after the search word(s) You can choose the collocation range by clicking two little boxes next to the COLLOCATE box.
- **POS LIST** (Parts of Speech List)

This function is used when you don't know the exact word (or a possible collocate of the word) you are looking for although you know which part of speech (noun, verb, adjective, etc.) you want to use. This is also called "wildcard" search. Place the cursor in the word or collocates box and click the drop-down arrow and select a part of speech that you are looking for. For example, if you don't know which preposition to use in a sentence like "I am going to New York ___ Spring Break", you may try the following:

The screenshot shows the COCA search interface. At the top, there is a 'DISPLAY' section with four radio buttons: LIST, CHART, KWIC (which is selected and highlighted in green), and COMPARE. Below this is the 'SEARCH STRING' section. It contains a 'WORD(S)' field with 'Spring Break' entered. The 'COLLOCATES' field has '[i*]' entered, with two small dropdown arrows to its right, one showing '1' and the other '0'. The 'POS LIST' field has a dropdown menu showing 'prep.ALL'. At the bottom left of the search string section is a 'RANDOM' button, and at the bottom right are 'SEARCH' and 'RESET' buttons.

From the search result, you can conclude that the preposition "on" is most frequently used with Spring break.

COCA (CORPUS OF CONTEMPORARY AMERICAN ENGLISH) EXERCISES

A. Learning collocations with COCA

Step 1: Look at the example sentence below. Answer questions a) - c). Then, circle the word that you should use in your paper.

I am fully / totally aware of the problem.

- In which genre is “totally” most frequently used?
- In which genre is “fully” most frequently used?
- So, which word would you use in your paper?

HINT) You should use CHART display option.

Step 2: Using COCA, find a better (more frequently used) collocate for the word “technology” in the sentence below.

I’m studying utilization / application of modern **technology** in classes.

HINT) You should use COMPARE display option with COLLOCATE search string.

Step 3: Look at the following sentences. Which one is correct? Use KWIC option in COCA to find out the answer.

- I am **looking forward to** meeting you in class.
- I am **looking forward to** meet you in class.

HINT) You should use KWIC display option.

Step 4: Look at the example sentences below. The underlined word in each sentence is an awkward collocate of the word in bold. Using COCA, find better collocates and revise the sentences. (You should keep the original meaning of each sentence) Decide which function you should use on your own. If you can’t think of which one to use, look at the hint below.)

- I hope to succeed the **goal**.
- There has been a hot **debate** over this issue.
- He firmly **recommended** this place.

HINT) You should use LIST display option with a wildcard (v*, adj*, adv*) COLLOCATE or a synonym COLLOCATE search string.

B. Be careful!

1. Always check the CONTEXT and GENRE! It is often dangerous to look at only the frequency count and decide which one to use. Having similar frequency counts does not always mean both words are possible in a given context. One limitation of COCA or some other corpora, however, is that we can usually only see “one line” of the context, which sometimes is not enough.
2. You may have multiple words in the [WORD] slot, but you cannot have strings of two or more words in the [COLLOCATES] box. Try reformulating your query so that the multiple words are in the [WORD] slot, and the single word is in the [COLLOCATES] box.
3. If there is no or few result showing, it happened for one of the following reasons:
 - One of the words could be spelled wrong, or an ungrammatical word.
 - The word combination is impossible or rare.
4. Currently, you cannot search for collocates of words that occur more than 3,000,000 times in the corpus, or which are both very frequent (sum of all forms) and have many different forms.
5. COCA can't do “Compare” words on 1- 2 letter words since it creates problems with script.