

Sample Corpus Worksheet

What are corpora?

- Corpora are a collection of data that can be used to look for patterns in language. There are different types: historical, American, British, Iweb(Internet Corpus), Now(gathers data every night from the internet), Harry Potter..... (You can make a corpus of anything).

Why should I care about corpora?

- It is sometimes difficult to understand how a word is used. Does the word have a negative connotation? When is the word used?
- Sometimes the language of non-native speakers can seem unnatural. How a word is used in a language is dependent on agreement of most speakers of the language (i.e., if a word is used incorrectly people will be confused)
- It is possible to have a sentence that is correct grammatically, but it will feel strange to a native speaker.
 - I want to **challenge** the test.
 - To a native speaker challenge as a verb is usually use with something more formal (e.g., court case, constitution, belief)
 - Challenge as a **verb** (means to contest, to fight against something believed to be wrong)
 - Challenge as a **noun** (means a difficult task)

COCA Corpus Basic functions:

- **Chart-** Shows where and when the word is used (Is this word academic?)
- **List-** Shows all occurrences of the word (concordance lines). (How is this word used in a sentence?)
- **Compare-** Shows the collocates of two words. (What is the difference between these two words? What words are usually used with each word?)

Sample Corpus Worksheet

- **Collocate-** Shows what is connected to one word. (What word is used with this word?) (The dark blue means it is a very strong association, lighter blue is less strong)

COLLOCATES **NATURALLY** **ADV** Advanced options [Collocates](#) [Clusters](#) [Topics](#) [Dictionary](#) [Texts](#) [KWIC](#)

+ NOUN	NEW WORD	?	+ ADJ	NEW WORD	?	+ VERB	NEW WORD	?	+ ADV	NEW WORD	?
85	3.42	substance	97	3.90	curious	1720	6.06	occur	342	2.96	quite
79	4.25	bacteria	93	5.45	inclined	223	2.66	produce	19	4.88	spontaneously
54	3.32	acid	64	4.20	gifted	219	3.24	assume	14	5.41	organically
53	3.23	chemical	51	5.82	curly	162	4.85	flow	13	3.17	ie
51	2.53	soil	45	2.94	organic	142	4.42	arise	13	4.86	effortlessly
51	3.63	compound	45	4.90	resistant	127	2.75	tend	9	3.73	artificially
47	2.03	hormone	30	2.03	immune	112	4.00	evolve	6	5.70	unusually

- **Key word in context- Shows** the word in sentence and highlights the different parts of speech (**noun** **pronoun** **proper** **noun** **adjective** **verb** **adverb** **preposition**)
- **Word-** Shows all the information about a word (e.g., collocates, clusters, concordance lines, definition, synonyms, links to translated sentences of the word)

*Note- POS= part of speech. Can help make your search more specific.



- The dark blue square shows where the word or phrase you entered will be. The scale on the left chooses how many words will be in front and the scale behind shows how many words you want to look at behind the word. This example has two green spaces highlighted so it will look at two words after.

Sketch Engine for Language Learners (SKELL)

Has similar functions to the COCA Corpus, but it is more useful for higher frequency words. Words that are more academic and less frequent show only concordance lines.

Functions in SKELL

- **Examples**
 - Can see examples of the word in context from concordance lines
 - These are a bit easier than COCA in my opinion

Sample Corpus Worksheet

- **Word Sketch**
 - Can be used to check for collocates and connotations
 - subject, object, modifiers, phrasal, etc.....
 - It is possible to see examples of word combinations
 - defense + force/ defense + out
- **Similar Words**
 - Can be used for synonyms and to see what words are usually commonly associated with it
 - Word cloud

Sample Corpus Worksheet

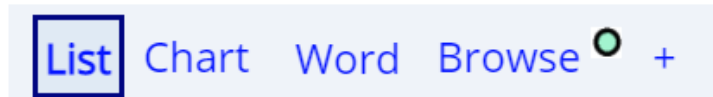
Practice

In groups you are going to use the Coca Corpus and SKELL to select the correct word or sentence for questions 1-5. **Make sure that you enter the information in the same way as the pictures.**

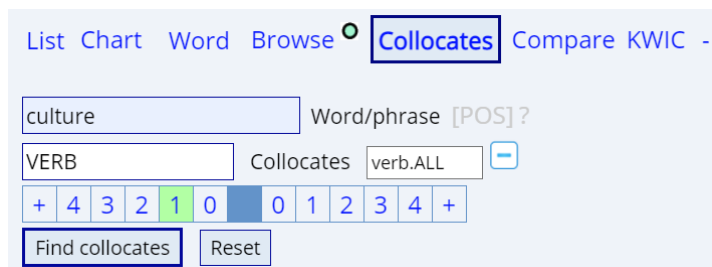
Part 1 : Use the Coca Corpus for questions 1- 3.

- <https://www.english-corpora.org/coca/>

1) **Push the + button and look for collocates.**



- Let's try searching for collocates (Word pairs)

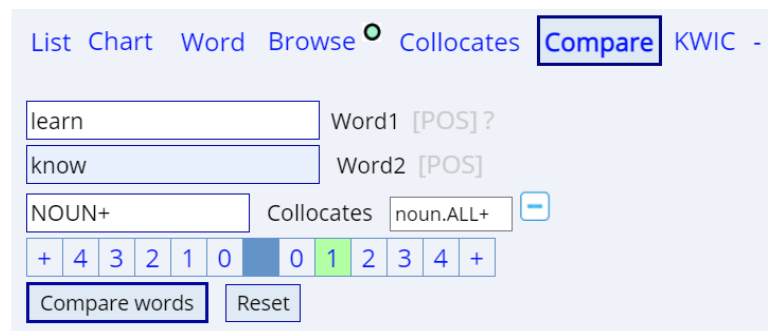


Search for collocates for the word "culture". The interface shows the word "culture" in the search box, with "Word/phrase [POS]?" below it. The "Collocates" dropdown is set to "verb.ALL". A bar chart shows the frequency of collocates, with the number "1" highlighted in green. The "Find collocates" button is highlighted.

- **Which is correct ?**

- touch culture
- experience culture

2) **Let's try a comparison search next**



Comparison search for the words "learn" and "know". The interface shows "learn" in the "Word1 [POS]?" box and "know" in the "Word2 [POS]?" box. The "Collocates" dropdown is set to "noun.ALL+". A bar chart shows the frequency of collocates, with the number "1" highlighted in green. The "Compare words" button is highlighted.

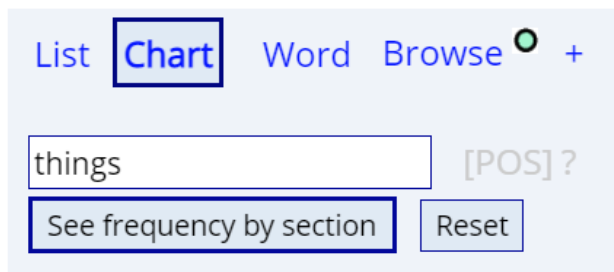
- **Which is correct?**

Sample Corpus Worksheet

- Students who are living overseas can know about the culture.
- Students who are living overseas can learn about the culture.

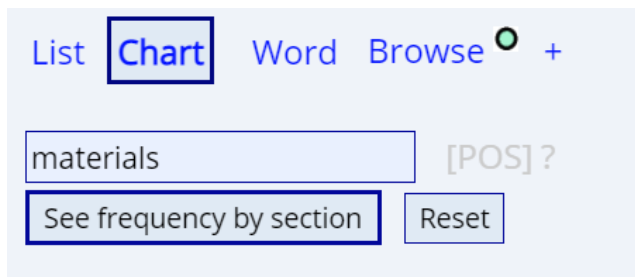
3) Next let's try using the chart function to see where a word is used.

- **First: Search for things**



The screenshot shows the SKELL interface with the 'Chart' tab selected. The search input field contains the word 'things'. Below the input field, there is a button labeled 'See frequency by section' and a 'Reset' button. The interface also shows navigation options: 'List', 'Chart', 'Word', 'Browse', and a plus sign.

- **Second: Search for materials**



The screenshot shows the SKELL interface with the 'Chart' tab selected. The search input field contains the word 'materials'. Below the input field, there is a button labeled 'See frequency by section' and a 'Reset' button. The interface also shows navigation options: 'List', 'Chart', 'Word', 'Browse', and a plus sign.

- **Which word is academic?**

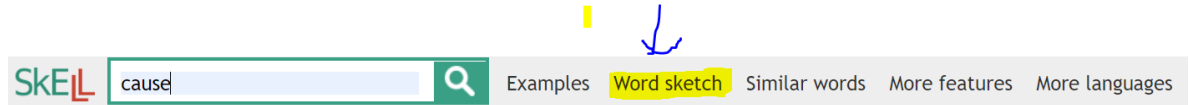
- things
- materials

Part 2: Use SKELL for questions 4- 5

Sample Corpus Worksheet

- <https://skell.sketchengine.co.uk/run.cgi/skell>

4) Let's try a collocation search using SKELL



- Type in cause
- Go to word sketch
- Look at **object of cause**

- **Which collocation is more likely to be used? Is cause used with positive words (+) or negative (-) words?**

- Cause great things to happen
- Cause damage

5) Let's try a collocation search using SKELL.



- Type the word mind in the search bar.
- Go to Word Sketch
- Look at the **adjectives with mind**

- **Which is correct?**

- I wanted to refresh my mind
- I wanted to clear my mind.