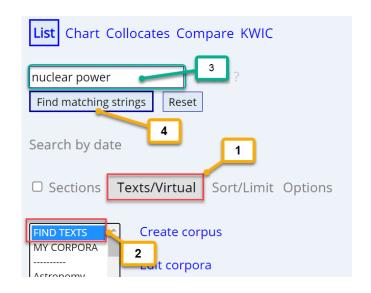
Creating and Using Virtual Corpora at English-Corpora.org (June 2021)

In 5-10 seconds, you can search through millions of texts to create "Virtual Corpora" (VC) on virtually any topic – *nuclear power, thyroid, vaccine, basketball, muggles, fake news, ransomeware,* etc. In 2-3 seconds more, you can create "keywords" from this corpus, and search for words, phrases, collocates, etc in the VC.

Creating Virtual Corpora



(From main Search page)

- 1. Click Select Texts/Virtual
- 2. Click Find texts
- 3. Enter the desired word or phrase
- 4. Click "Find matching strings"

On the following "results" page



1. Click Save List and then 2. Name the list

(Optional) 3. Choose the number of texts in the VC, or (de)select specific texts A. See information on each of the texts, including relevance

(Optional) SAVE AS: nuclear OR ADD TO:SELECT SUBMIT RESET										
HELP	100	TEXT 2	4	# WORDS	# HITS ‡	RELEVANCE \$	PE			
1		3 11-12-15: COUNTERCURRENTS.ORG			66	6,740.9				
2	2	US 13-02-11: EXTREMETECH	29288	49	A 1,673.0					
3	~	US 19-08-05: MARGINAL REVOLUTION	10584	49	4,629.6					
4	~	US 20-03-11: REUTERS	1947	46	23,626.1					
5	2	ALL 12-03-10: LINKS INTERNATIONAL		3180	∆ 1	12 893 1				

Using and searching your Virtual Corpora

You can see a list of all of the Virtual Corpora you have created (potentially hundreds or thousands of texts). Just one corpus is shown in the screenshot below.



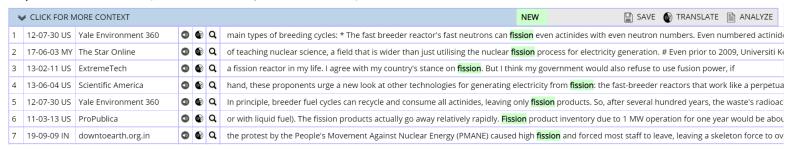
- 1. Click on a part of speech to see Keywords from the Virtual Corpus
- 2. See the number of texts and the number of words in each Virtual Corpus
- 3. Delete corpora, temporarily deactivate them, rename then, move/add texts between corpora, etc.

In 1-2 seconds, you can see a list of hundreds of keywords in your Virtual Corpus, ranked by relevance.

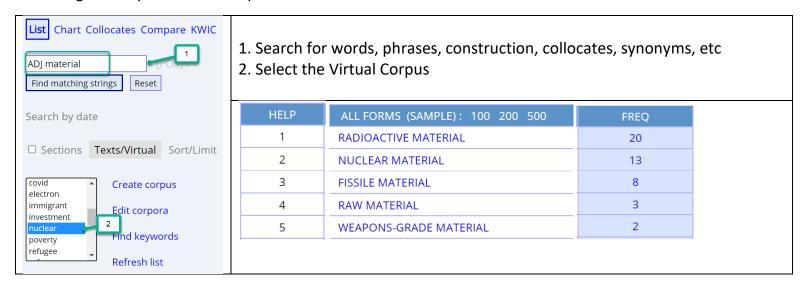
NUCLEAR	[375,166 WORDS, 100 TEXTS] (3.0% OF TOTAL)	NOUN VERB ADJ ADV N+N ADJ+N			[ALL CORPORA] SAVE LIST		
HELP	WORD (CLICK FOR CONTEXT)	FREQ	# TEXTS	FREQ 45 10 TEXTS	ENTIRE CORPUS	EXPECTED	
1	REACTOR	1395	87	1.7 3	28,609	845.1	
2	THORIUM	69	17	1.4	1,623	47.9	
3	PLUTONIUM 1	142	24	1.4	3,505	103.5	
4	FISSION	62	21	1.0	2,018	59.6	
5	RADIOACTIVITY	57	24	0.9	2,066	61.0	
_	KILIVE	40	47	0.7	2.254	CC F	

- 1. Click to see the words in context, within your Virtual Corpus (see below)
- 2. Information on the number of hits and number of texts in the VC that have the word
- 3. Adjust the "specificity" of the keywords; lower Frequency or Texts for more specific words, increase for more general, and then click 'Specific".

Keyword in Context (from list of Keywords, above)



Searching within your Virtual Corpora



You can also compare the frequency of words and phrases in the different Virtual Corpora that you have created. And you can also use metadata for the texts (source, title of text, date, author, genre, sub-genre, etc) as you create the Virtual Corpora. Click here for more details.

With the corpora from English-Corpora.org, you can create Virtual Corpora on virtually any topic within 4-5 seconds, and then search within that corpus or find keywords in 2-3 seconds more – even with corpora containing billions of words and tens of millions of texts.