Task: COCA demonstration

Screenshot 1

Corpus of Contemporary Americ... × | G hydroxychloroquine - Google Se... × | +

← → C 🔒 english-corpora.org/coca/

Apps M Gmail ▶ YouTube CodeVIT

Corpus of Contemporary American English

SEARCH | CHART | CONTEXT | OVERVIEW

List Chart Word Browse +

hydroxychloroquine   [POS]

Find matching strings   Reset

☐ Sections Texts/Virtual Sort/Limit Options

🏠 (HIDE HELP)   NOT LOGGED IN

Try the new **Coronavirus Corpus**: hundreds of millions of words of data, and updated daily.

ⓘ Download the corpus (and corpus-based frequency data) for offline use

The Corpus of Contemporary American English (COCA) is the only large, genre-balanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English.

The corpus contains more than one billion words of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the update in March 2020): TV and Movies subtitles, blogs, and other web pages.

Click on any of the links in the search form to the left for context-sensitive help, and to see the range of queries that the corpus offers.

There are three main ways to search the corpus:

First, you can browse a frequency list of the top 60,000 words in the corpus, including searches by word form, part of speech, ranges in the 60,000 word list.

09:47 PM

Screenshot 2

Corpus of Contemporary Americ... × | G hydroxychloroquine - Google Se... × | +

← → C 🔒 english-corpora.org/coca/

Apps M Gmail ▶ YouTube CodeVIT

Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | OVERVIEW

(SHUFFLE)

CLICK FOR MORE CONTEXT   ☐ [?]   SHOW DUPLICATES

1	2019	ACAD	Urological Science	A B C	basal level of autophagy. 12 Inhibition of basal autophagic activities through treatment with chloroquine or hydroxychloroquine, two well-known autophagy i
2	2019	ACAD	Urological Science	A B C	, Yang SC, Tsai TF, Chen HE, et al. Chloroquine and hydroxychloroquine inhibit bladder cancer cell growth by targeting basal autophagy and enhancing apopto
3	2017	ACAD	Clinical Therapeutics	A B C	and novel biologic agents, compared with conventional immunosuppressive agents (eg, prednisone, hydroxychloroquine, azathioprine, cyclophosphamide, m
4	2017	ACAD	Clinical Therapeutics	A B C	enrolled adults with moderate to severe SLE despite use of corticosteroids, antimalarial agents (hydroxychloroquine), and additional immunosuppressive age
6	2017	ACAD	Clinical Therapeutics	A B C	).67 # Very-low-quality evidence suggests that background therapy with prednisone alone or combined hydroxychloroquine plus adjunctive anti-interferon- r
7	2017	ACAD	Clinical Therapeutics	A B C	interferon alfa receptor antibody Immunosuppressive agents: corticosteroid and conventional immunosuppressive agents (e.g., Hydroxychloroquine, Azathic
8	2017	ACAD	Clinical Therapeutics	A B C	Adults with inactive SLE on stable doses of azathioprine (<= 3 mg/kg/day), hydroxychloroquine (<= 400 mg/day), chloroquine (<= 250 mg/day), and/ or
9	2017	ACAD	Clinical Therapeutics	A B C	low No difference GRADE Summary of Findings Appendix D Table VI Prednisone alone or combined hydroxychloroquine plus adjunctive rontalizumab in adul
10	2017	ACAD	Clinical Therapeutics	A B C	have had an inadequate response to immunosuppressive agentsSetting: OutpatientIntervention: Prednisone alone or combined hydroxychloroquine plus ad
11	2015	ACAD	ACPJrnlClub	A B C	were compared with methotrexate in 26 trials, placebo in 8 trials, sulfasalazine plus hydroxychloroquine plus methotrexate in 2 trials, sulfasalazine in 1 trial,
12	2014	ACAD	PracticeNurse	A B C	recommends that all patients should be commenced on three DMARDs (methotrexate, sulphasalazine and hydroxychloroquine) with monthly reviews to mon
13	2014	ACAD	PracticeNurse	A B C	or long term warfarin if the patient has already experienced a DVT or PE.24 # Hydroxychloroquine is an important drug therapy for SLE, unless side effects pi
14	2014	ACAD	PracticeNurse	A B C	20:206 # 25. Tang C, Godfrey T, Stawell R, Nikpour M. Hydroxychloroquine in lupus: emerging evidence supporting multiple beneficial effects. Internal Medici
15	2014	ACAD	PracticeNurse	A B C	cardiac manifestations of neonatal lupus in fetuses of mothers with systemic lupus erythematosus exposed to hydroxychloroquine. Annals of Rheumatic Dise
16	2014	ACAD	ACPJrnlClub	A B C	patients (about 50%), oral triple therapy (MTX, sulfasalazine, and hydroxychloroquine) yields similar clinical outcomes to combination therapy with MTX plus a
17	2012	WEB	niams.nih.gov	A B C	whether estrogen differentially regulates target organs. Pregnancy # Breakthroughs # The discovery of hydroxychloroquine's safety during pregnancy and th

09:48 PM

Corpus of Contemporary American English

| SEARCH | FREQUENCY | CONTEXT | OVERVIEW |

List Chart Word Browse Collocates **Compare** KWIC -

recent | Word1 [POS]
latest | Word2 [POS]
* | Collocates [POS]

| + | 4 | 3 | 2 | 1 | 0 | | 0 | 1 | 2 | 3 | 4 | + |

Compare words | Reset

☐ Sections Texts/Virtual Sort/Limit Options

(HIDE HELP) NOT LOGGED IN

COMPARE WORDS display

Compare the collocates of two words, to see how they differ in meaning and usage. For example, utter and sheer (note the negative collocates with *utter*), warm and hot, small and little, or adjectives near boy and girl.

By comparing collocates, you can move far beyond the simplistic entries in a thesaurus, to "tease out" slight differences in words, or (as in the case of *boy* and *girl*) what is the difference in what is being said about two different things.

Please review the discussion of collocates to see how to select the span for the collocates.

Corpus of Contemporary American English

| SEARCH | FREQUENCY | CONTEXT | OVERVIEW |

SEE CONTEXT: CLICK ON NUMBERS (WORD 1 OR 2)
SORTED BY FREQUENCY: CHANGE TO RATIO

WORD 1 (W1): **RECENT** (2.33)

	WORD	W1	W2	W1/W2	SCORE
1	THE	65094	58250	1.1	0.5
2	.	60834	23492	2.6	1.1
3	IN	60541	11837	5.1	2.2
4	,	56988	22278	2.6	1.1
5	A	40760	5004	8.1	3.5
6	OF	33365	12744	2.6	1.1
7	YEARS	19239	193	99.7	42.8
8	TO	17788	9901	1.8	0.8
9	AND	17621	8687	2.0	0.9
10	MOST	15467	384	40.3	17.3
11	THAT	13367	3838	3.5	1.5
12	"	10851	5474	2.0	0.9
13	ON	10382	6919	1.5	0.6
14	'S	10223	9001	1.1	0.5
15	HAVE	8629	1554	5.6	2.4
16	BY	8159	1539	5.3	2.3
17	HAS	8126	1372	5.9	2.5

WORD 2 (W2): **LATEST** (0.43)

	WORD	W2	W1	W2/W1	SCORE
1	THE	58250	65094	0.9	2.1
2	.	23492	60834	0.4	0.9
3	,	22278	56988	0.4	0.9
4	OF	12744	33365	0.4	0.9
5	IN	11837	60541	0.2	0.5
6	TO	9901	17788	0.6	1.3
7	'S	9001	10223	0.9	2.1
8	AND	8687	17621	0.5	1.1
9	IS	7433	7036	1.1	2.5
10	ON	6919	10382	0.7	1.6
11	"	5474	10851	0.5	1.2
12	A	5004	40760	0.1	0.3
13	HIS	4927	5682	0.9	2.0
14	THIS	4570	4161	1.1	2.6
15	WITH	4554	7150	0.6	1.5
16	FOR	4386	6053	0.7	1.7
17	FROM	4260	5956	0.7	

[HELP...]

Screenshot 1

Corpus of Contemporary American English

| SEARCH | FREQUENCY | CONTEXT | OVERVIEW |

List **Chart** Word Browse +

`michael phelps` [POS]

See frequency by section   Reset

☐ Sections Texts/Virtual Sort/Limit Options

(HIDE HELP)   NOT LOGGED IN

CHART display

If you are interested in a set of words or a grammatical construction, then the LIST option shows the frequency of each matching form (*end up being, ended up saying*, etc), while the CHART option shows the total frequency in each section. (in COCA, the genres and five year blocks). It is also possible to see the frequency of words and phrases by sub-genre and/or year, e.g. Mitt Romney or Obama.

For example, some cases of lexical change over time: morph, old school, freak out, think outside the box, throw someone under the bus, be likely a|the.

Some cases of lexical variation between genres: I guess, , you know ,, muffled, frowned, validity, correlate, . in particular, a lot of, kind of NOUN, type of NOUN.

Of course you can also search for grammatical constructions, such as : and I'm like, get passive, end up V-ing, be passive, appear to V, must + verb.

Screenshot 2

Corpus of Contemporary American English

| SEARCH | CHART | CONTEXT | OVERVIEW |

CHANGE TO VERTICAL CHART / CLICK TO SEE CONTEXT

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	524	68	79	13	137	1	70	147	9	1	0	54	145	70	107
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	0.53	0.53	0.64	0.10	1.09	0.01	0.56	1.21	0.08	0.01	0.00	0.37	1.00	0.48	0.74
SEE ALL SUB-SECTIONS AT ONCE															

Task: CFG for other language and loaned words in mother tongue

Q) 4 loaned words in your mother toungue

मराठी (Marathi)   English Meaning

बाग (Baag)   Garden

मुजरा (Mujra)   Mujra   } Taken from Persian Words

दरबार (Darbar)   Court

जकात (Jakat)   Tax

मुलाखत (Mulakhat) Interview

आजारी (Aazari)   Sick

b) CFG for own language (or other than english)

```
              S
          ____/\____
         NP          VP
        /  \         /  \
      PRP   NN     VBG    V
       |     |      |     |
      मी    गाणे   ऐकत   आहे
```

(I am listening song)

Task: Does Porter stemmer use any lexicon?

No, Porter Stemmer doesn't require any lexicon. The Porter stemming algorithm (or 'Porter stemmer') is a process for removing the commoner morphological and inflexional endings from words in English. Its main use is as part of a term normalisation process that is usually done when setting up Information Retrieval systems.