COCA: Corpus of Contemporary American English

1. What is COCA:

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, and even by pronunciation. This should be particularly useful for language learners and teachers.

Second, you can search by individual word, and see collocates, topics, clusters, websites, concordance lines, and related words for each of these words. Note that some of these searches are unique to COCA and iWeb.

Third, you can search for phrases and strings. And because the corpus is optimized for speed, searches for substrings (*ism, un*able) and phrases are very fast, e.g.: got VERB-ed, BUY * ADJ NOUN, "gorgeous" NOUN -- and even high frequency phrases like: from ADJ to ADJ, phrasal verbs, or NOUN NOUN.

You might pay special attention to the comparisons between genres and years and virtual corpora, which allow you to create personalized collections of texts related to a particular area of interest.

2. Searching on COCA: Harry Potter

List Chart Word Browse +

```
harry potter                    [POS]

Find matching strings    Reset
```

☐ Sections Texts/Virtual Sort/Limit Options

3. Context:
   Context shows various columns like years, sites, categories, type of platform and the sentences where the words are encountered.

					CLICK FOR MORE CONTEXT	☐ [?]	SHOW DUPLICATES
1	2012	BLOG	ontheredcarpet.com	A B C	Biblical hero who builds an ark. The Darren Aronofsky-directed movie also stars " **Harry Potter** " actress Emma Watson, Jennifer Connelly and Anthony Ho	
2	2012	BLOG	aintitcoolnews.com	A B C	not even get started about a grown ass man cumming in his pants about **Harry Potter**.... and im not bagging on HP.... but when a grown ass man in	
4	2012	BLOG	womackcompany.com	A B C	adopted and orphaned children, from Moses to Cinderella to Oliver Twist to **Harry Potter**. So why do many parentless children feel compelled to hide the	
5	2012	BLOG	...micbookresources.com	A B C	to think that DC are looking at the guy as their own answer to **Harry Potter** or some other mass market magical selling point. # And the previous column a	
6	2012	BLOG	...micbookresources.com	A B C	London? What sort of backlash would that have unleashed? What if the **Harry Potter** films were set in Los Angeles, and used American actors? What if the	
7	2012	BLOG	...micbookresources.com	A B C	# But Constantine being British IS important, just as it's important with **Harry Potter** and James Bond, and just as it's important that Spider-Man's America	
8	2012	BLOG	...micbookresources.com	A B C	the 2005 film realize how many adventure heroes have British origins? Aside from **Harry Potter**: # if you look at the supposedly well-known adventure fra	
9	2012	BLOG	laura-joakimson.com	A B C	how to climb into books like Lucy slipping into a psychedelic wardrobe, or **Harry Potter** standing on Platform 9. Ask any kid which matters more, a good,	
10	2012	BLOG	korrektivpress.com	A B C	text-messaging and hyperactive pop culture metabolism, are more enchanted by wide-eyed, quidditch-playing **Harry Potter** of Hogwarts than by the smir	
11	2012	BLOG	5newsonline.com	A B C	the University of Arkansas -- Fort Smith, and Oklahoma Baptist University. # **Harry Potter** fans, you are in fact reading this correctly, these people will be p	
12	2012	BLOG	5newsonline.com	A B C	# The sport is played very similar to the way it's shown in **Harry Potter**, just minus that whole detail of flight. Players are required to have brooms	
13	2012	BLOG	5newsonline.com	A B C	at all times. There are seven players to a team, like in **Harry Potter**, three chasers, two beaters, one keeper, and of course a seeker	
14	2012	BLOG	...randparentsvoice.com	A B C	by Sarah Addison Allen # Girl With a Pearl Earring by Tracy Chevalier # **Harry Potter** Books by J. K. Rowling # Ireland by Edward Rutherford # Kathy Reich	
15	2012	BLOG	aintitcool.com	A B C	of course, as Stephen King said (I'm paraphrasing), " **Harry Potter** is about friendship, loyalty, courage and doing the right thing. Twilight is	
17	2012	BLOG	aintitcool.com	A B C	day. That day has not come yet. Same thing with the **Harry Potter** books and movies. The more I hear or read about them the less interest	
19	2012	BLOG	...history.umwblogs.org	A B C	liked which included health tips and a bunch of nerdy things related to **Harry Potter** and my current literary love, The Hunger Games. # Overall, this week	
20	2012	BLOG	vigilantcitizen.com	A B C	one day. # You can also see the Unicorn on the cover of **Harry Potter** and the Sorceres Stone -- again this is more occult brainwashing for children -- pavir	
21	2012	BLOG	hollywoodlife.com	A B C	come with terms of acknowledging being an actor at all. When I did **Harry Potter**, I remember looking at Dan Radcliffe, Rupert Grint, and Emma Watson a	
22	2012	BLOG	journalgazette.net	A B C	replica of a 95-liter turbo diesel engine; or Diagon Alley, from? **Harry Potter**?; or an 8-foot long Rube Goldberg machine, kind of like a detailed	
23	2012	BLOG	patheos.com	A B C	not believe as we do. # 3. My Church never officially condemned **Harry Potter**. On the contrary, it recognizes that the arts have a rightful place in	

4. Change/Compare:
   This compares how many times the words were encountered on different corpora

Re-do last search	Corpus (click to use)	Description (corpora arranged by size)
✓	iWeb	14 billion words from the Web
✓	NOW	102 billion, Web news, 2010 - yesterday
✓	GloWbE	1.9 billion, Web, 20 countries
✓	Wikipedia	1.9 billion, Wikipedia
✓	Hansard	1.6 billion, British Parliament
✓	COCA	1.0 billion, US, 1990-2015
✓	EEBO	755 million, 1470s-1690s
✓	COHA	400 million, US, 1810s-2000s
✓	NEW Coronavirus	370 million words; current through yesterday
✓	TV	325 million words in 75,000 very informal TV shows
✓	Movies	200 million words in 25,000 very informal movies
✓	Supreme Court	130 million words, 1790s-present
✓	TIME	100 million, US, 1923-2006
✓	SOAP	100 million, US, 1990s-2000s
✓	BNC	100 million, British, 1980s-1993
✓	CORE	50 million words, Web genres
✓	Strathy	50 million, Canada