

英语语料库(English-Corpora.org)使用指南

[Mark Davies](#), 语言学教授

2020 年 11 月

为何语言变体如此重要

词频

词组和搭配 (和型式)

语法 / 句法

语义 (通过搭配词体现的词义和用法)

历时变体 (近期变化)

方言变体

虚拟语料库 (聚焦某一特定话题)

为语言学习者和教师准备的工具

其它工具及功能

English-Corpora.org 是世界范围内使用最为广泛的语料库集 (高度可检索文本集)。本语料库每月的使用人数超过 13 万, 遍布 140 多个国家。此外, 全球有几百所大学持有**学术研究许可证**, 用户凭此证可获取高级访问权限。

在本语料库集的基础上, **成千上万篇学术文章**、论文以及学位论文得以完成, 而这些文本也支撑起世界范围内涵盖各个层次的语言及语言学课程。近五至十年里, 几乎所有关于“用语料库进行英语教学”的书籍都围绕本语料库集展开 (因其创建于杨百翰大学, 有时也被称为“杨百翰大学系列语料库”)。

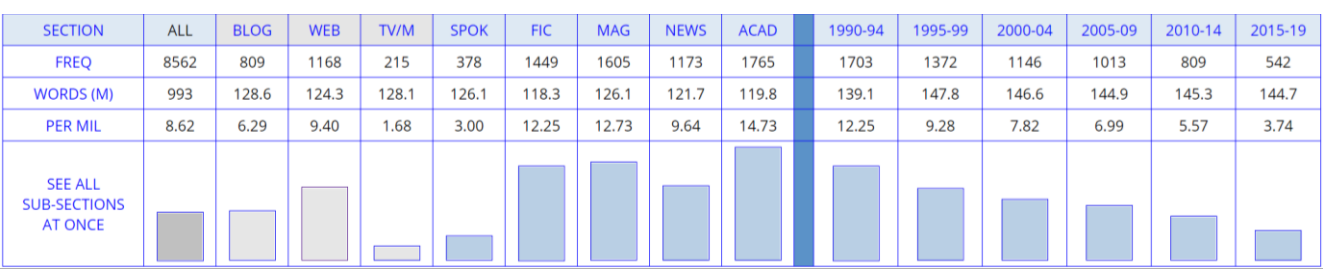
自 2005 年第一批语料库发布以来, 现在共建有 17 个语料库:

	Corpus	# words	Dialect	Time period	Genre(s)
1	iWeb: The Intelligent Web-based Corpus	14 billion	6 countries	2017	Web
2	News on the Web (NOW)	11.3 billion+	20 countries	2010-yesterday	Web: News
3	Global Web-Based English (GloWbE)	1.9 billion	20 countries	2012-13	Web (incl blogs)
4	Wikipedia Corpus	1.9 billion	(Various)	2014	Wikipedia
5	Hansard Corpus	1.6 billion	British	1803-2005	Parliament
6	Corpus of Contemporary American English (COCA)	1.0 billion	American	1990-2019	Balanced
7	Early English Books Online	755 million	British	1470s-1690s	(Various)
8	Coronavirus Corpus	673 million+	20 countries	2020-yesterday	Web: News
9	Corpus of Historical American English (COHA)	400 million	American	1810-2009	Balanced
10	The TV Corpus	325 million	6 countries	1950-2018	TV shows
11	The Movie Corpus	200 million	6 countries	1930-2018	Movies
12	Corpus of US Supreme Court Opinions	130 million	American	1790s-present	Legal opinions
13	Corpus of American Soap Operas	100 million	American	2001-2012	TV shows
14	British National Corpus (BNC)	100 million	British	1980s-1993	Balanced
15	TIME Magazine Corpus	100 million	American	1923-2006	Magazine
16	Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s	Balanced
17	CORE Corpus	50 million	6 countries	2014	Web

为何语言变体（如此）重要 (返回开头)

本语料库集尤为关注（不同体裁、历史时期和方言间）的英语变体，使得它有别于其它任何语料库。那些语料库仅提供大“团”数据，很少提及变体。变体为何重要呢？让我们以单词 *seldom* 为例，在 COCA (含 10 亿词的当代美式英语语料库，后文简称 COCA) 中，该词在正式场景中的使用频数比在非正式场景中使用的频数更高，并且其使用频数正随时间推移而骤降。

(注意: 在示例 *seldom* 和本文件中其它的此类示例中，单击蓝色链接进行检索。在浏览器上，可以通过“在新标签页上进行搜索”并随后关闭来使检索更加便捷。)



FIND SAMPLE: 100 200 500 1000
PAGE: << < 1 / 86 > >>

CLICK FOR MORE CONTEXT				<input type="checkbox"/> [?]	SAVE LIST	CHOOSE LIST	CREATE NEW LIST	<input type="checkbox"/> [?]	SHOW DUPLICATES
1	1996	ACAD	Bioscience	A B C	marina may reflect dispersal mechanism constraints. Although both species are perennial, <i>C. demersum</i> seldom produces seeds and disperses largely b				
2	1998	NEWS	CSMonitor	A B C	Rams have a small (30,000 seat) stadium, which works out because it seldom is filled. Admits media-relations director Gary Ozello, "We've been down so				
3	1999	ACAD	AnthropolQ	A B C	government's intention to use village courts to reinforce traditional means of dispute settlement was seldom evident in Kwanga courts. There the magist				
4	2012	WEB	...irdworldtraveler.com	A B C	mar the opinions of multi-member tribunals. But the process was professional in a way seldom achieved in military courts, and the records and judgmer				
5	2002	SPOK	NPR_ATC	A B C	result, some of the most influential and important figures in politics are people you seldom hear about in campaign news reports. They are the shoo-ins.				
6	2012	BLOG	...ncebasedmedicine.org	A B C	of the " atypical " pneumonia such as Mycoplasma or Clamydia pneumonia, which are seldom so severe as to cause death, would have been expected to				
7	2012	BLOG	dailykos.com	A B C	of tomorrow. The disappointment that greets us can be overpowering when our dreams so seldom meet reality. Do not let this become your governing f				
8	1996	MAG	AmSpect	A B C	the anchors and commentators who were superfluous. As noted by many observers, they seldom knew when to shut up. # Sometimes, of course, this w				
9	2002	ACAD	EnvironHealth	A B C	primary reason that stools are not often tested for enteric virus is that there is seldom a benefit to the patient. # The high fraction of asymptomatic infec				
10	2018	MAG	MarketWatch	A B C	have the reputation of being particularly risky, but the statistics show that they were seldom the worst performer in any of these time periods. # If you're				
11	1991	FIC	BkSF:HeirtoEmpire	A B C	He'd seen a lot of marketplaces on a lot of different planets, but seldom one so crowded. Crowded with more than just locals, too. Scattered throughout				
12	2012	WEB	...info.library.unt.edu	A B C	the electorate identify as key issues. In the years before September 11, terrorism seldom registered as important. To the extent that terrorism did break				

如果一个大型在线语料库只能告诉你，*seldom* 在 170 亿个单词中出现了 8 万 7 千次，那么这个语料库的用处并不大。因为学生们永远都不会知道，他们在用这个词的时候给人感觉其实要么像个七八十岁老头，要么太过郑重其事，或者两者兼有。这只是个检索频数的简单例子。而对于无数单词（频数、语义和用法）以及语法结构来说也是这样，只知道其频数还不够。语言变体十分重要，而 English-Corpora.org 是唯一一个能详细展现英语变体的语料库。

词频 (返回开头)

首先介绍最基本的功能：用户可以检索任何单词或短语在本语料库集不同版块出现的频数，也可以在子版块中检索频数（仅在某些语料库中）。例如，单词 *strategic* 在 COCA 中的学术板块中的频数最多，而在整个学术版块中，该词又在商业、历史和 法律/政治学几个子版块中的频数最多。

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD		1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	26103	3198	3435	459	2753	422	3614	3502	8720		3812	2547	3503	3027	3461	3120
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	26.29	24.87	27.65	3.58	21.83	3.57	28.66	28.77	72.79		27.41	17.24	23.90	20.88	23.83	21.56
SEE ALL SUB-SECTIONS AT ONCE																

History	Education	Geog/SocSci	Law/PolSci	Humanities	Phil/Rel	Sci/Tech	Medicine	Misc	Business
2552	1197	1130	1526	346	185	533	658	167	401
13.4	15.8	20.0	12.3	16.2	7.8	17.5	10.8	4.8	1.2
190.51	75.88	56.42	124.21	21.35	23.59	30.54	60.87	34.66	339.77

用户可以检索任何单词、词组或子字符串（例如 [words with *break*](#)，表示检索包含子字符串 *break* 的单词），并在语料库的各版块中查看所有匹配结果。比如，COCA 会依次呈现检索结果在博客、其它网页、电视/电影字幕、无脚本电视和广播话语类节目、小说、杂志、报纸和学术期刊中的频数。

HELP	<input type="checkbox"/>	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/>	BREAK	145708	13941	13958	26949	43525	14183	14284	13326	5542
2	<input type="checkbox"/>	BREAKING	41693	5553	5197	5555	6947	5410	5308	5406	2317
3	<input type="checkbox"/>	BREAKFAST	33610	2751	3022	7074	2027	7859	5456	4806	615
4	<input type="checkbox"/>	BREAKS	24620	3745	3861	2305	2750	3182	3772	3343	1662
5	<input type="checkbox"/>	BREAKDOWN	9321	1376	1371	745	1011	615	1363	1159	1681
6	<input type="checkbox"/>	OUTBREAK	7711	572	910	421	926	243	1133	1149	2357
7	<input type="checkbox"/>	BREAKTHROUGH	6998	653	769	534	1176	294	1748	1191	633
8	<input type="checkbox"/>	BREAKUP	4170	387	441	455	488	305	946	646	502
9	<input type="checkbox"/>	OUTBREAKS	3990	244	456	42	212	59	574	413	1990
10	<input type="checkbox"/>	HEARTBREAKING	3161	514	536	180	717	215	474	481	44
11	<input type="checkbox"/>	GROUND BREAKING	2967	394	562	144	259	63	662	592	291
12	<input type="checkbox"/>	BREAKTHROUGHS	2197	232	280	98	253	60	713	289	272
13	<input type="checkbox"/>	BREAKOUT	2287	389	334	97	246	43	511	569	98
14	<input type="checkbox"/>	HEARTBREAK	2171	271	336	292	283	269	359	314	47

用户也可以在语料库中任意两个版块间进行比较，比如通过对比发现，特别不正式的电视/电影字幕（下左）和相当正式的学术文本（下右）中，包含子字符串 *break* 的单词出现的次数要多得多。

SEC 1 (TV/MOVIES): 128,074,534 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	BREAKIN	172	0	1.3	0.0	134.3
2	HEARTBREAKER	101	1	0.8	0.0	94.5
3	DEAL-BREAKER	35	1	0.3	0.0	32.7
4	JAILBREAK	41	0	0.3	0.0	32.0
5	BREAK-DANCE	20	1	0.2	0.0	18.7
6	HEARTBREAKERS	55	3	0.4	0.0	17.1
7	BED-AND-BREAKFAST	45	3	0.4	0.0	14.0
8	BREAK-IN	403	35	3.1	0.3	10.8
9	BREAKFAST	7074	615	55.2	5.1	10.8
10	LATE-BREAKING	30	4	0.2	0.0	7.0
11	BREAKER	403	56	3.1	0.5	6.7
12	TIEBREAKER	34	5	0.3	0.0	6.4
13	BREAK-DANCING	20	3	0.2	0.0	6.2
14	HEARTBREAK	292	47	2.3	0.4	5.8

SEC 2 (ACADEMIC): 119,790,456 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	BREAKEVEN	62	1	0.5	0.0	66.3
2	OUTBREAKS	1990	42	16.6	0.3	50.7
3	PATH-BREAKING	41	1	0.3	0.0	43.8
4	BREAKPOINT	32	1	0.3	0.0	34.2
5	PATHBREAKING	38	0	0.3	0.0	31.7
6	STRIKEBREAKING	34	0	0.3	0.0	28.4
7	RULE-BREAKING	26	1	0.2	0.0	27.8
8	BREAKPOINTS	20	0	0.2	0.0	16.7
9	ICEBREAKERS	20	3	0.2	0.0	7.1
10	BREAKAGE	97	15	0.8	0.1	6.9
11	OUTBREAK	2357	421	19.7	3.3	6.0
12	STRIKEBREAKERS	42	9	0.4	0.1	5.0
13	BREAKDOWNS	171	44	1.4	0.3	4.2
14	BREAK-EVEN	31	9	0.3	0.1	3.7

研究人员也可以查看在某一语料（或子语料）中比在另一语料中出现频数高得多的所有单词。例如，下左图显示的单词是在 [COCA: Academic: Medicine](#)（学术语料中的医学子语料）中的高频词，而非 COCA:

Academic（学术体裁）中普遍使用的。这样，用户就可以轻松找到如商业、医学、法律或工程等相关领域的词汇。

SEC 1 (ACAD:Medicine): 10,809,528 WORDS							SEC 2 (ACADEMIC): 108,980,928 WORDS						
	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	MASTOID	226	1	20.9	0.0	2,278.5	1	GIFTED	8038	8	73.8	0.7	99.7
2	PAROTID	388	2	35.9	0.0	1,955.9	2	THEOLOGICAL	3626	4	33.3	0.4	89.9
3	TONSILLAR	184	0	17.0	0.0	1,702.2	3	FEMINIST	2605	3	23.9	0.3	86.1
4	MEDIASTINAL	142	0	13.1	0.0	1,313.7	4	ISLAMIC	7551	10	69.3	0.9	74.9
5	TRANSCUTANEOUS	122	1	11.3	0.0	1,230.0	5	ARAB	8591	12	78.8	1.1	71.0
6	SCAPULAR	114	1	10.5	0.0	1,149.3	6	FICTIONAL	2853	4	26.2	0.4	70.7
7	PLEOMORPHIC	110	1	10.2	0.0	1,109.0	7	PEDAGOGICAL	2079	3	19.1	0.3	68.7
8	OTOLOGIC	110	1	10.2	0.0	1,109.0	8	LITERARY	9567	14	87.8	1.3	67.8
9	FASCIAL	118	0	10.9	0.0	1,091.6	9	PROTESTANT	1880	3	17.3	0.3	62.2
10	ANTIHYPERTENSIVE	115	0	10.6	0.0	1,063.9	10	RULING	1760	3	16.1	0.3	58.2
11	OTOTOXIC	114	0	10.5	0.0	1,054.6	11	RITUAL	2342	4	21.5	0.4	58.1
12	ETHMOID	113	0	10.5	0.0	1,045.4	12	IRAQI	3439	6	31.6	0.6	56.9
13	SPHENOID	112	0	10.4	0.0	1,036.1	13	BIBLICAL	2060	4	18.9	0.4	51.1
14	COELIAC	110	0	10.2	0.0	1,017.6	14	NATIONALIST	2044	4	18.8	0.4	50.7

词组与搭配（词串） [\(返回开头\)](#)

当然，用户不止可以检索单个单词。下图展示了在 COCA 里，词组 *soft + NOUN* 在不同体裁中的结果，请注意学术领域的 *soft tissue(s)*, *power*, *skills*, 电视/电影的 *soft spot*, 小说的 *soft voice*, *light*, *skin*, *touch*, *music*, 和新闻与杂志的 *soft drink(s)* 或 *landing* 的结果。所以，一个语料库，即使库容“巨大”（150-200 亿单词），如果没有区分体裁，就无法呈现上述的差异结果。

HELP	<input type="checkbox"/>	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/>	SOFT TISSUE	1120	62	67	74	39	36	100	35	707
2	<input type="checkbox"/>	SOFT DRINKS	1109	123	123	42	90	83	304	296	48
3	<input type="checkbox"/>	SOFT MONEY	790	21	34	12	446	8	78	153	38
4	<input type="checkbox"/>	SOFT SPOT	867	133	110	166	63	159	135	86	15
5	<input type="checkbox"/>	SOFT DRINK	721	43	60	48	77	68	191	199	35
6	<input type="checkbox"/>	SOFT VOICE	546	12	39	10	11	351	53	52	18
7	<input type="checkbox"/>	SOFT POWER	421	49	53	1	64	1	51	32	170
8	<input type="checkbox"/>	SOFT TISSUES	328	13	23	5	6	10	41	6	224
9	<input type="checkbox"/>	SOFT LANDING	274	42	19	18	38	17	56	63	21
10	<input type="checkbox"/>	SOFT LIGHT	247	22	25	9	1	121	38	21	10
11	<input type="checkbox"/>	SOFT PEAKS	216	2	4		1	2	144	62	1
12	<input type="checkbox"/>	SOFT SKIN	207	9	24	28	6	114	19	4	3
13	<input type="checkbox"/>	SOFT TOUCH	213	22	26	28	15	43	51	25	3
14	<input type="checkbox"/>	SOFT SKILLS	219	41	41		7		13	13	104
15	<input type="checkbox"/>	SOFT MUSIC	176	7	27	39	5	48	27	15	8

用户可以通过比较语料库内两种不同的体裁，找出一种体裁中使用频数更高的词组。例如，以下截图是 *phrasal verbs with out*（含有 *out* 的动词词组）分别在小说（左）和学术（右）中的高频词组。

SEC 1 (FICTION): 118,322,084 WORDS							SEC 2 (ACADEMIC): 119,790,456 WORDS						
	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	STARED OUT	950	3	8.0	0.0	320.6	1	CONTRACTING OUT	34	1	0.3	0.0	33.6
2	GLANCED OUT	230	1	1.9	0.0	232.9	2	CARDED OUT	63	2	0.5	0.0	31.1
3	STEPS OUT	490	3	4.1	0.0	165.4	3	PARTIALED OUT	20	0	0.2	0.0	16.7
4	LEANING OUT	139	1	1.2	0.0	140.7	4	COOLING OUT	32	2	0.3	0.0	15.8
5	FLUNG OUT	119	1	1.0	0.0	120.5	5	BEARS OUT	33	3	0.3	0.0	10.9
6	LETS OUT	113	1	1.0	0.0	114.4	6	PHASING OUT	52	5	0.4	0.0	10.3
7	SHOOK OUT	205	2	1.7	0.0	103.8	7	CARRIED OUT	4467	434	37.3	3.7	10.2
8	STEPPED OUT	1426	14	12.1	0.1	103.1	8	SINGLES OUT	68	7	0.6	0.1	9.6
9	LAUGHED OUT	300	3	2.5	0.0	101.2	9	OPT OUT	56	6	0.5	0.1	9.2
10	PEERED OUT	369	4	3.1	0.0	93.4	10	POINTS OUT	2826	306	23.6	2.6	9.1
11	WHIPS OUT	83	1	0.7	0.0	84.0	11	BORNE OUT	216	25	1.8	0.2	8.5
12	WANDERED OUT	83	1	0.7	0.0	84.0	12	BEAR OUT	57	7	0.5	0.1	8.0

型式 (返回开头)

语料库还可以显示单词和短语的搭配。单词并非是孤立出现的，学习者需了解一个特定单词的搭配型式。例如，从下图可看到，[作动词的 account](#) 后面几乎总是跟着 *for*：

49	1991	MAG	Sierra	A	B	C	# Asbestos mining in Canada provides scarcely 2,000 jobs , and accounts for only one percent of total mineral exports. Substitutes are
50	2012	BLOG	...ealclearpolitics.com	A	B	C	capita GDP : # As you can see , once you account for population growth , we are still struggling to get
51	2012	BLOG	randomhouse.com	A	B	C	Americans in the first three decades of the 20th century . Account for Roosevelt 's position . Taylor argues that the struggle
52	2015	ACAD	SchoolPsych	A	B	C	accuracy and fluency factors, although kindergarten WRF accounted for somewhat more variance (43% to 54%) in the prediction
53	2012	BLOG	loadedboards.com	A	B	C	we had to explore means of analyzing the presser blank to account for springback of the process by which the severity of a deck
54	2005	MAG	Sportsill	A	B	C	people do n't know that. " Yes, what exactly accounts for that difference , the black and the blue ? Well,
55	2005	ACAD	Environment	A	B	C	agricultural production and crop yields but neglected to account for the additional downstream benefits that better land-use
56	2004	SPOK	NPR_Morning	A	B	C	Mr-DAVIS : We have felt like all along that nobody has accounted for the death of 160 people. This is the first time
57	1999	ACAD	IBMR&D	A	B	C	. Results of step 5 are used with closed-form equations to account for the delay impact of noise on nets with minimal timing slac
58	2015	ACAD	LangSpeechHearing	A	B	C	, therefore, did not seem likely that utterance length would account for the difference in failure rates between sets. # A second
59	2012	BLOG	...bellinghamherald.com	A	B	C	hunch that these polls of " likely voters " are n't accounting for the enthusiasm gap . Fewer Dems will show up at the

而 [fathom](#) 前几乎总有一个否定词。这便是为什么对母语者而言，*I totally fathom what you're saying* 这样的句子（*fathom* 前无任何否定）听起来很奇怪。

29	2001	MAG	Redbook	A	B	C	not in prison watches senior golf. I still ca n't fathom why anyone watches hydroplane racing. The jump rope
30	2012	WEB	open.salon.com	A	B	C	and irrational fear causes them to hate what they ca n't fathom . We humans have always done this, in various settings
31	2012	MAG	Prevention	A	B	C	each side. # Shop smarter # If you ca n't fathom going through a flip-flop-less summer, opt for a more
32	2012	FIC	Bk:KingsBlood	A	B	C	your judgment. " # For some reason Dawson could n't fathom , the blush in Jorey 's cheeks returned and deepened. His
33	2015	MAG	MotherEarth	A	B	C	denial. In the 18th century, most people could n't fathom that any creature that had once lived on Earth could have
34	1999	NEWS	Chicago	A	B	C	captivated by a top-notch thriller. # Many readers could n't fathom that these men were aboard submarines packed with
35	2012	FIC	SouthernRev	A	B	C	had been caught in a crime, though she could n't fathom what she was guilty of. # " I should have known
36	2012	BLOG	socialmediatoday.com	A	B	C	tweet with the hashtag #apowerfulnoise in it anywhere and NCM Fathom will donate 10 cents for it, up to 50,000 Tweets.
37	2018	FIC	Windsor Review	A	B	C	clues very seriously. # She does things she would never fathom doing : hide things ! At a dinner party, she turns
38	2014	FIC	Bk:HeritageCyador	A	B	C	after five years of trying, for reasons he can not fathom he has been unable to create shields directly linked to himself,
39	2012	WEB	forums.adobe.com	A	B	C	which has a folder icon for a reason I can not fathom), pick " Edit UV Properties " to get the dialog
40	2016	TV	Underground	A	B	C	just let you, did n't he ? I can not fathom men 's disregard for their children. They are only concerned
41	2012	BLOG	politics.gather.com	A	B	C	set these men up to die deliberately. I can not fathom as to why they would do this, But then I am
42	2007	FIC	FantasySciFi	A	B	C	It seems you feel bound by some compulsion I can not fathom to honor me with your presence and with the company of your
43	2012	BLOG	...logs.mercurynews.com	A	B	C	the post and shanking a second kick, I can not fathom why Chip Kelly would n't put Rob Beard (who was handling

在展示单词的搭配型式方面，语料库远胜于简单的词典。

语法/句法 (返回开头)

语料库的最佳用途之一是研究句法结构的频数和用法。例如，试看 "like 结构"（*and I'm like, he can't do it*, 或者 *but she was like, let's just buy it*），语料库可显示[所有匹配短语](#)的频数，以及它们在语料库[各体裁中的频数](#)（下图显示的是 "like" 结构在 COCA 中各体裁和 1990-2019 年各时间段的频数）。

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019
1	<input type="checkbox"/> AND I WAS LIKE,	1795	73	74	569	807	16	161	88	7	31	75	157	293	432	660
2	<input type="checkbox"/> AND I'M LIKE,	1190	51	41	408	463	35	129	58	5	15	68	108	214	301	392
3	<input type="checkbox"/> AND HE 'S LIKE,	456	18	7	190	191	6	30	13	1	4	19	41	84	125	158
4	<input type="checkbox"/> AND IT 'S LIKE,	414	17	15	93	216	6	32	30	5	22	50	44	43	94	129
5	<input type="checkbox"/> AND YOU 'RE LIKE,	411	23	11	154	161	3	36	21	2	4	14	22	50	123	164
6	<input type="checkbox"/> AND THEY 'RE LIKE,	335	12	10	121	148	2	24	15	3	3	26	26	51	77	130
7	<input type="checkbox"/> AND HE WAS LIKE,	325	12	16	97	131	2	41	24	2	5	19	25	46	78	124
8	<input type="checkbox"/> AND SHE 'S LIKE,	262	6	4	85	128	13	18	6	2	3	8	30	52	68	91
9	<input type="checkbox"/> AND IT WAS LIKE,	198	4	3	44	98	2	29	18		7	30	20	35	49	50
10	<input type="checkbox"/> AND SHE WAS LIKE,	190	13	9	43	89	8	21	6	1	1	9	23	23	54	58
11	<input type="checkbox"/> AND THEY WERE LIKE,	166	13	6	35	81	3	19	9		2	7	18	26	30	64
12	<input type="checkbox"/> AND WE 'RE LIKE,	99	3	3	34	44		8	7		1	6	13	17	22	34

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	7270	329	263	2257	3156	126	699	394	46	140	393	639	1145	1780	2581
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	7.32	2.56	2.12	17.62	25.02	1.06	5.54	3.24	0.38	1.01	2.66	4.36	7.90	12.25	17.83
SEE ALL SUB-SECTIONS AT ONCE															

CLICK FOR MORE CONTEXT					<input type="checkbox"/> [?]	<div>SAVE LIST</div>	<div>CHOOSE LIST</div>	<div>-----</div>	<div>CREATE NEW LIST</div>	<div></div> [?]	SHOW DUPLICATES
1	2018	SPOK	CBS_Morning	A B C	you get into the game, you want to play some more and they're like , well, if you buy this then you play more, you get a						
2	2002	MOV	An Evening with Kevin Smith	A B C	vault. " I was like, " For what? " And she's like , " I don't know. " I was like, " Is it						
3	2016	SPOK	ABC: The View	A B C	because if your parents show body confidence, if you have that and you're like , and you're like, this is what I have, this is what						
4	2014	SPOK	CNN: CNN Live Event	A B C	" Why couldn't you be normal and just be gay. And I was like , " Mom, who said that? " UNIDENTIFIED-FEMALE# I need a strong man						
5	2012	MOV	Sleepwalk with Me	A B C	, I should close strong. What Spanish do I know? And I'm like , " I know. I'll say, Long live the Immigrant. "						
6	1993	SPOK	PBS_NewsHour	A B C	about the trees, and I'm going to show you. And I was like , hey, you don't have to show me nothin', but what						
7	1993	MOV	...ve! The Valentine's Day Massacre	A B C	passport is gone. Yeah. This bird came in and... And I was like , " Huh? " You don't want to meet my family. I						
8	2001	SPOK	ABC_GMA	A B C	training command, and my training command took care of it, and it was like , 'We're not going to have this,' and it stopped.						
9	2018	SPOK	ABC_20/20	A B C	got one video that's coming up on a million views. And it's like , wow, you know, they all want to hear what I have to						
10	2018	SPOK	NPR_AskMe	A B C	West Florissant, and these four officers come up to me. And they're like , hey, you can't stand there. I was like, I just						
11	2019	SPOK	NPR_ATCW	A B C	, and I was - we were searching for a title. And I was like , well, how about " Room 41? " I mean, that's						
12	2002	MOV	An Evening with Kevin Smith	A B C	was like, " I'm here to interview you. " And I was like , " Get out of here. You? " I couldn't not talk						
13	2011	SPOK	CBS_48Hours	A B C	that one of his clients had put up for adoption. BRUCE-LISKER: And she was like , what? Was sort of thrown, but came to just love it.						
14	2008	SPOK	NBC_Dateline	A B C	I was just talking to my dad about it one day. And I was like , Dad, I don't -- I don't get this. Why me						

我们再看看 COCA 中 "**BE passive**" (如 *he was hired; it was paid*) 或 "**GET passive**" (如 *he got hired; it got paid*) 的频数。BE 被动态在正式体裁中出现的频数更高 (这就反驳了被动态主要用于 "草率" 讲话中的观点), 但随着时间的推移略有减少。而 GET 被动态则更多地出现在非正式体裁中, 其频数随着时间的推移增加。因此, 若有人用英语写一篇学术论文, 使用 BE 被动态要比使用 GET 被动态恰当得多, 因为后者太不正式。

BE + V-ed	SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
	FREQ	3301114	661282	850189	115423	279587	148822	252117	321244	672450	329769	306317	293036	272612	271486	316423
	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	3,324.31	5,141.63	6,842.36	901.22	2,216.56	1,257.77	1,999.48	2,638.73	5,613.55	2,371.43	2,072.87	1,999.30	1,880.79	1,868.94	2,186.14
	SEE ALL SUB-SECTIONS AT ONCE															
GET + V-ed	SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
	FREQ	208091	34353	26869	50926	33009	21756	19638	17673	3867	20394	23604	24436	25483	26234	26718
	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	209.55	267.10	216.24	397.63	261.69	183.87	155.74	145.17	32.28	146.66	159.73	166.72	175.81	180.60	184.59
	SEE ALL SUB-SECTIONS AT ONCE															

由于 COCA 是唯一具有体裁多样、库容大、语料新等特点的英语语料库, 数以百计类似的英语句法变异方面的深度研究都是基于它进行的。

语义 (词义和用法) ([返回开头](#))

遵循 "由单词的搭配可知其词" 这一理念, **搭配词** (邻近词) 对深入了解一个单词或短语的 **语义和用法** 极其有用。在 iWeb (由来自网络的 140 亿单词组成) 和 COCA (10 亿词、体裁均衡) 语料库中, 用户可看到按词性划分的搭配词频数 (并显示搭配词更倾向于出现在检索词的前或后, 以及两词的 "紧密度")。例如, 这些是 iWeb 语料库中 *hormone* 的搭配词 (先点 "WORD" 搜索, 再点 "COLLOCATES" 查看搭配词):

COLLOCATES				Advanced options				Collocates				Clusters Topics Dictionary Websites KWIC			
HORMONE				NOUN											
+ NOUN		NEW WORD	?	+ ADJ		NEW WORD	?	+ VERB		NEW WORD	?	+ ADV		NEW WORD	?
26921	4.45	level		21212	9.97	thyroid		14705	4.64	produce		2121	4.24	naturally	
24776	5.89	growth		5215	5.77	male		9157	4.31	release		196	3.92	genetically	
14859	6.48	therapy		5184	3.40	human		6836	3.19	cause		148	3.06	negatively	
13529	3.80	body		4967	5.29	female		5592	6.46	regulate		115	4.15	chemically	
12996	6.29	stress		4366	3.17	natural		4433	2.90	increase		109	4.77	biologically	
10128	4.63	production		3592	9.13	steroid		4313	3.52	affect		103	3.94	artificially	
8575	5.51	sex		3176	8.40	adrenal		4160	9.33	secrete		64	2.75	adversely	
8548	5.87	replacement		3085	7.97	stimulating		3642	5.40	balance		55	2.85	orally	
7912	9.81	cortisol		2833	6.20	synthetic		3186	6.38	stimulate		53	3.70	chronically	
6696	8.11	testosterone		2457	10.15	parathyroid		3019	3.13	control		53	3.72	abnormally	
5912	8.61	estrogen		2309	2.78	normal		1766	12.43	luteinizing		52	3.55	structurally	
4951	7.12	insulin		2274	2.86	responsible		1478	4.24	trigger		47	6.74	superfamily	
4859	6.79	antibiotic		2226	9.19	pituitary		1356	3.78	decrease		44	7.46	acromegaly	
4830	3.62	blood		1710	8.27	anabolic		1213	3.25	bind		43	2.77	selectively	
4588	8.08	imbalance		1618	11.92	bioidentical		1160	5.31	disrupt		38	6.06	synthetically	

搭配词通常是看 "邻近"的词（如左边四个词到右边四个词）。话题（English-Corpora.org 特有的功能）可查看语篇中任何地方共现的词。在许多情况下，话题有助于更好地了解一个词的词义和用法（再看看 iWeb 语料库中 *hormone* 的例子）：

TOPICS (more)

[symptom](#), [blood](#), [diet](#), [stress](#), [gland](#), [muscle](#), [fat](#), [testosterone](#), [body](#), [estrogen](#), [pregnancy](#), [protein](#), [cell](#), [supplement](#), [disease](#), [tissue](#), [treatment](#), [doctor](#), [vitamin](#), [acid](#)

COLLOCATES (more)

NOUN [level](#), [growth](#), [therapy](#), [body](#), [stress](#), [production](#), [sex](#), [replacement](#)

VERB [produce](#), [release](#), [cause](#), [regulate](#), [increase](#), [affect](#), [secrete](#), [balance](#)

ADJ [thyroid](#), [male](#), [human](#), [female](#), [natural](#), [steroid](#), [adrenal](#), [stimulating](#)

ADV [naturally](#), [genetically](#), [negatively](#), [chemically](#), [biologically](#), [artificially](#), [adversely](#), [orally](#)

搭配词有时反映出一个词具有与最初预期不同的 "语义韵"，"语义韵"反映某些词的搭配词更多是积极还是消极色彩。例如，我们可看到在 COCA 中（作为动词的）[cause](#) 的名词搭配词是多么消极：

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	PROBLEMS	4720	831	894	295	728	172	745	600	455
2	DAMAGE	3919	608	743	255	501	143	709	492	468
3	PAIN	2187	314	342	268	250	239	422	179	173
4	TROUBLE	2048	224	219	480	267	400	204	166	88
5	PEOPLE	2050	440	412	99	396	74	262	231	136
6	DEATH	1770	223	312	171	303	122	190	243	206
7	HARM	1928	489	393	65	196	78	221	175	311
8	CANCER	1430	130	220	99	228	32	436	170	115
9	DISEASE	1194	107	178	24	150	10	343	110	272
10	LOT	1171	210	142	151	386	55	98	101	28
11	PROBLEM	1158	207	183	108	273	40	165	87	95
12	LOSS	969	164	160	41	70	13	182	104	235
13	INJURY	804	97	167	31	66	26	125	113	179
14	DEATHS	774	126	147	32	106	18	114	147	84
15	CONCERN	683	66	79	19	125	30	124	107	133

搭配词也可用来研究近义词间的区别，如对比 [totally](#) 和 [completely](#) (+ADJ)，可观察到 *totally* 的搭配词更为非正式（左）。

WORD 1 (W1): TOTALLY (0.65)

	WORD	W1	W2	W1/W2	SCORE
1	CUTE	35	0	70.0	108.4
2	FUN	42	1	42.0	65.0
3	HOT	107	4	26.8	41.4
4	GREAT	51	2	25.5	39.5
5	GAY	47	5	9.4	14.6
6	LAME	51	6	8.5	13.2
7	AWESOME	276	33	8.4	13.0
8	CREEPY	24	3	8.0	12.4
9	EXCITED	23	3	7.7	11.9
10	EXCELLENT	21	3	7.0	10.8
11	SWEET	24	4	6.0	9.3
12	COOL	375	75	5.0	7.7

WORD 2 (W2): COMPLETELY (1.55)

	WORD	W2	W1	W2/W1	SCORE
1	CONTROLLABLE	21	1	21.0	13.6
2	RANDOMIZED	24	2	12.0	7.7
3	BARE	56	7	8.0	5.2
4	IMMOBILE	30	4	7.5	4.8
5	REVERSIBLE	30	4	7.5	4.8
6	UNNOTICED	30	4	7.5	4.8
7	RED	22	3	7.3	4.7
8	DRY	212	33	6.4	4.1
9	IDENTICAL	24	4	6.0	3.9
10	MAD	99	18	5.5	3.6
11	STILL	22	4	5.5	3.6
12	UNUSABLE	22	4	5.5	3.6

单词语义和用法也会因体裁而不同。例如，试看 [care 的搭配词](#) 在小说中（左图；搭配词更多是接受照顾的对象）和在学术文本中（右图；搭配词更多是提供照顾的主体）的不同：

SEC 1 (FICTION): 118,322,084 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	DAD	36	2	0.3	0.0	18.2
2	HORSES	17	1	0.1	0.0	17.2
3	AUNT	17	1	0.1	0.0	17.2
4	THING	32	2	0.3	0.0	16.2
5	NIGHT	30	2	0.3	0.0	15.2
6	DOG	40	3	0.3	0.0	13.5
7	DADDY	13	1	0.1	0.0	13.2
8	MOM	38	3	0.3	0.0	12.8
9	KITCHEN	12	1	0.1	0.0	12.1
10	GARDEN	12	1	0.1	0.0	12.1
11	GRANDMA	11	1	0.1	0.0	11.1
12	TOWN	11	1	0.1	0.0	11.1

SEC 2 (ACADEMIC): 119,790,456 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	SETTINGS	399	1	3.3	0.0	394.1
2	MODEL	236	1	2.0	0.0	233.1
3	SUPPORT	205	1	1.7	0.0	202.5
4	COSTS	339	2	2.8	0.0	167.4
5	PROVIDERS	779	5	6.5	0.0	153.9
6	GROUP	155	1	1.3	0.0	153.1
7	PRACTICE	309	2	2.6	0.0	152.6
8	SYSTEMS	298	2	2.5	0.0	147.2
9	PHYSICIANS	272	2	2.3	0.0	134.3
10	MEMBERS	126	1	1.1	0.0	124.5
11	INDIVIDUALS	123	1	1.0	0.0	121.5
12	SERVICES	1048	9	8.7	0.1	115.0

搭配词也可超越严格的 "语义"，反映出对于不同话题“我们在谈论什么”。例如，来看看 [Asia](#)（左图；更侧重国家和机构）和 [Africa](#)（右图；更侧重个人、健康和福祉）的搭配词。

WORD 1 (W1): ASIA (0.50)

	WORD	W1	W2	W1/W2	SCORE
1	COOPERATION	66	11	6.0	12.0
2	SUMMIT	84	16	5.3	10.5
3	ECONOMIES	124	25	5.0	9.9
4	MARKETS	177	40	4.4	8.9
5	STABILITY	83	19	4.4	8.8
6	RADIO	51	14	3.6	7.3
7	INFLUENCE	64	26	2.5	4.9
8	SOCIETY	155	66	2.3	4.7
9	FOUNDATION	66	29	2.3	4.6
10	PRESENCE	80	40	2.0	4.0
11	SECURITY	102	59	1.7	3.5
12	MARKET	64	39	1.6	3.3

WORD 2 (W2): AFRICA (2.00)

	WORD	W2	W1	W2/W1	SCORE
1	AIDS	286	14	20.4	10.2
2	AID	162	10	16.2	8.1
3	COAST	429	41	10.5	5.2
4	ARTS	101	12	8.4	4.2
5	LIFE	125	17	7.4	3.7
6	HUMANS	77	11	7.0	3.5
7	CHILDREN	197	29	6.8	3.4
8	EDUCATION	75	12	6.3	3.1
9	WORK	117	19	6.2	3.1
10	CONTINENT	257	43	6.0	3.0
11	HEALTH	71	12	5.9	3.0
12	WOMEN	188	32	5.9	2.9

English-Corpora.org 中的语料库是唯一可按同义词进行检索的语料库，这意味着可以把检索集中在意义和形式（多词）上。这对非母语者而言非常有用，因为他们可以从多个 "相互竞争" 的词中看到哪些词在特定语境中被实际使用（如 ["strong" argument](#)），从而让他们的文章或语言更 "接近母语者"。

ListChartWordBrowse+

=strong ARGUMENT

[POS]

Find matching strings

Reset

☒ Sections

Texts/Virtual

Sort

HELP	<input type="checkbox"/>	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/>	STRONG ARGUMENT	331	83	57	3	54	8	38	25	63
2	<input type="checkbox"/>	CONVINCING ARGUMENT	218	55	44	12	16	13	21	23	34
3	<input type="checkbox"/>	POWERFUL ARGUMENT	148	19	20	2	28	4	17	17	41
4	<input type="checkbox"/>	PERSUASIVE ARGUMENT	137	21	23	12	16	5	15	14	31
5	<input type="checkbox"/>	EFFECTIVE ARGUMENT	39	6	7	2	12		5	2	5
6	<input type="checkbox"/>	POTENT ARGUMENT	12	1	4		2		2	2	1
7	<input type="checkbox"/>	FORCEFUL ARGUMENT	13	3	4		1		1	1	3
8	<input type="checkbox"/>	VIGOROUS ARGUMENT	10		2	1			1		6
9	<input type="checkbox"/>	INFLUENTIAL ARGUMENT	7		1				1		5

同义词也会因体裁而变化。例如，在小说（左）和学术语篇（右）中检索 [strong 的同义词](#)。要在同义词

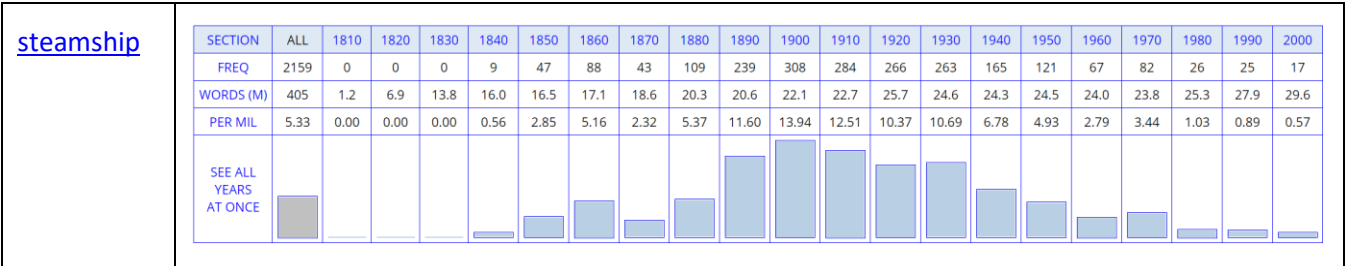
词典中找到所有这些词并不难，但只有语料库可以显示一个词的同义词会因体裁而异，如在小说体裁中，作家可能会用 *beefy*, *burly*, *strapping lumberjacks* 作为“strong”的同义替代，而在学术写作中与“strong”同义的则是 *effective*, *compelling*, *persuasive arguments*。

SEC 1 (FICTION): 118,322,084 WORDS							SEC 2 (ACADEMIC): 119,790,456 WORDS						
	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	BEEFY	301	7	2.5	0.1	43.5	1	EFFECTIVE	28807	1272	240.5	10.8	22.4
2	BURLY	650	27	5.5	0.2	24.4	2	ROBUST	2829	444	23.6	3.8	6.3
3	STRAPPING	297	21	2.5	0.2	14.3	3	DEEP-SEATED	260	49	2.2	0.4	5.2
4	SPICY	507	53	4.3	0.4	9.7	4	COMPELLING	2845	602	23.7	5.1	4.7
5	PUNGENT	575	70	4.9	0.6	8.3	5	PERSUASIVE	1360	298	11.4	2.5	4.5
6	BITING	1545	230	13.1	1.9	6.8	6	CLEAR-CUT	405	92	3.4	0.8	4.3
7	BRIGHT	16050	2542	135.6	21.2	6.4	7	DURABLE	683	160	5.7	1.4	4.2
8	STURDY	1369	240	11.6	2.0	5.8	8	DEDICATED	3496	1166	29.2	9.9	3.0
9	HOT	21731	3877	183.7	32.4	5.7	9	ZEALOUS	217	81	1.8	0.7	2.6
10	GLARING	1326	247	11.2	2.1	5.4	10	RESILIENT	550	210	4.6	1.8	2.6
11	DAZZLING	857	215	7.2	1.8	4.0	11	POTENT	1149	444	9.6	3.8	2.6
12	STOUT	953	270	8.1	2.3	3.6	12	POWERFUL	11539	5884	96.3	49.7	1.9

历时变体 (返回开头)

English-Corpora.org 中的许多语料库提供了非常有用的语言演变数据，比如聚焦 15 世纪-17 世纪（在线版早期英语书籍语料库 EEBO），1810-2009 年（历时美国英语语料库 COHA），1800-2018 年（美国最高法院语料库 US Supreme Court），1803-2003 年（英国议会议事录语料库 Hansard; British Parliament），或是涵盖 1926-2006 年（美国时代杂志语料库 TIME）。电影语料库（Movies）（1930-2010 年）和电视（TV）语料库（1950-2010 年）是仅有的两个大型语料库，提供大量相当非正式化口语变化数据。而研究人员也可在如 COCA（1990-2019）、在线新闻语料库（NOW）（2010-2020）和冠状病毒语料库（Coronavirus）（2020）几个库中关注近期的语言变化。后两者每晚都会更新数百万词的数据。总的来说，这些语料库有数十亿词，库容大多是同类历时语料库的 50-100 倍，这使得研究人员能观察到更为广泛的现象。此外，它们提供的检索范围比在 Google Books n-grams 中简单精确检索词和短语要广得多。

在入门层面上，研究人员可以查看单词和短语在各个年代出现的频数。例如，以下是 COHA 语料库（含 4 亿单词，涵盖 1810-2009 年）的截图，可以看到 *steamship* 一词在各个年代出现的频数，以及 *Reds* 在各个年代甚至每年出现的频数（注意 1953 年，即美国参议院举行数次麦卡锡听证会那年）。像搜索 *a most ADJ NOUN* 一样，研究人员也可以查找短语，包括词性。



REDS

SECTION	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	2496	0	0	1	4	3	7	26	14	36	54	65	179	202	191	567	315	125	337	166	204
WORDS (M)	405	1.2	6.9	13.8	16.0	16.5	17.1	18.6	20.3	20.6	22.1	22.7	25.7	24.6	24.3	24.5	24.0	23.8	25.3	27.9	29.6
PER MIL	6.16	0.00	0.00	0.07	0.25	0.18	0.41	1.40	0.69	1.75	2.44	2.86	6.98	8.21	7.84	23.10	13.14	5.25	13.31	5.94	6.90
SEE ALL YEARS AT ONCE																					
1950		1951		1952		1953		1954		1955		1956		1957		1958		1959			
88		92		61		123		64		22		28		15		25		49			
2.6		2.5		2.5		2.4		2.3		2.5		2.6		2.3		2.7		2.4			
33.26		36.74		24.60		52.12		27.92		8.83		10.79		6.49		9.36		20.15			

CLICK FOR MORE CONTEXT

☐ [?]

SAVE LIST

CHOOSE LIST

CREATE NEW LIST

[?]

SHOW DUPLICATES

1

1953

MAG

Time

A

B

C

, though abundant, is sluggish in following a moving target. # If the Reds have good reasons for attacking at night, the U.N. has equally good ones

2

1953

FIC

ReturnLannyBudd

A

B

C

eyes open. It's happening all the time; just a short time ago the Reds took away half a dozen students from the university. It caused an uproar,

3

1951

MAG

Time

A

B

C

especially since U.N. forces in the central mountains were bravely and skillfully holding the Reds back from mountain passes that meant access to the plains and

4

1952

MAG

Time

A

B

C

presumed safety. Subverted by agents, most of their Chinese crews defected to the Reds. They grabbed eleven of the planes and took off for Mao's mainland.

5

1952

MAG

Time

A

B

C

Premier Huy Kanthoul was more interested in plaguing the French than in keeping out the Reds. # Last week the King decided to take matters into his own hands

6

1951

MAG

Time

A

B

C

was printed a terse " Count your men. " # This week the Reds broke contact over most of a 70-mile front, fell back to lick their wounds

7

1953

NEWS

Chicago

A

B

C

finger, and wrote in blood, " The Communists never defeated us. " Reds Are Disorderly When the second convoy of American trucks with Red prisoners passed

8

1956

MAG

ReadersDigest

A

B

C

to the cover of an abutment on the far side of the stream. The Reds were on a low hill, 50 yards away. Every few minutes Page or

9

1954

NF

HowColor-TuneYour

A

B

C

in your room. If you don't care for an overstimulating effect, avoid reds and provide for yourself a background of light delicate tones derived from yellow or orange

10

1951

NEWS

Chicago

A

B

C

R. E. Libby, a Ridgway negotiator at the truce talks, told the Reds: " Your prisoners of war tell us they saw large numbers of United Nations

a most

ADJ NOUN

SECTION	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	11087	45	376	802	857	908	762	817	935	805	929	789	729	483	346	362	277	336	254	138	137
WORDS (M)	405	1.2	6.9	13.8	16.0	16.5	17.1	18.6	20.3	20.6	22.1	22.7	25.7	24.6	24.3	24.5	24.0	23.8	25.3	27.9	29.6
PER MIL	27.38	38.10	54.28	58.22	53.40	55.13	44.68	44.01	46.02	39.08	42.04	34.76	28.42	19.63	14.21	14.75	11.55	14.11	10.03	4.94	4.63
SEE ALL YEARS AT ONCE																					

CLICK FOR MORE CONTEXT

☐ [?]

SAVE LIST

CHOOSE LIST

CREATE NEW LIST

[?]

SHOW DUPLICATES

1

1887

MAG

Atlantic

A

B

C

week // later, on the 18th, it ratified the Constitution unanimously. A most auspicious beginning had thus been made. Three States, one third of the wh

2

1880

MAG

Atlantic

A

B

C

does not speak a word of). These formalities settled, I mounted a most ungainly mule, and preceded by a train of others, bearing instruments and prov

3

1889

NF

Arena Volume4

A

B

C

all the musical work necessary in the plays of that time. She was a most attractive member of the company, and as Morgiana (Forty Thieves), Lucy

4

1887

FIC

SamanthaAtSaratoga

A

B

C

her high-heeled shoes. They wuz both dressed up perfectly beautiful, and made a most splendid show. Wall, they went into a store on their way to the

5

1886

FIC

MillMystery

A

B

C

house was, as far as I could judge from the exterior, of a most respectable character, and the lady who answered my somewhat impatient summons w

6

1883

FIC

GuardianAngel

A

B

C

, his true destiny was the glorious career of a poet. It was a most pleasing circumstance, that his mother, while she fully recognized the propriety of his

7

1887

MAG

Atlantic

A

B

C

ingenious work to me, before I had thought of visiting England, was a most gratifying circumstance. I have mentioned the hospitalities extended to me

8

1889

NF

ChopinOtherMusical

A

B

C

I am neither a patriotic Frenchman nor a consumptive Pole, and I am a most ardent admirer of Schumann; nevertheless I uphold my former opinion, a

9

1885

MAG

Century

A

B

C

Charles de Kay, is conspicuous for height of aim, and certainly for a most resolute purpose. In these days it is bracing to see a man of his

10

1883

MAG

NorthAmRev

A

B

C

prove powerless even though we were members. Men voted for delegates and substitutes*with a most absurd ignorance of what they might do. Until

11

1883

MAG

Atlantic

A

B

C

, I visited Irving's grave, in the crypt of the cathedral, a most dismal place, and was touched to see the bronze tablet that marked its site

12

1889

FIC

WhoSpokeNext

A

B

C

. He was always a man of cordial friendliness, and he now expressed a most gratifying interest when I told him what I was going to do in Boston.

研究人员可以找到所有年代中符合要求的字符串的频数，例如可以在 COHA 语料库中查看*[*ism words](#)*（以 *ism* 结尾的单词）在不同年代的频数。可以看到，*patriotism*、*despotism*，和 *heroism* 在 19 世纪出现的频数更高，*socialism*、*communism*，以及 *nationalism* 在 20 世纪中期出现频数更高，*capitalism* 和 *terrorism* 在 20 世纪末、21 世纪初出现频数更高。

HELP	CONTEXT	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
1	<input type="checkbox"/> CRITICISM	13510	25	156	244	341	370	408	690	682	706	1016	1212	1123	860	771	922	974	975	835	659	541
2	<input type="checkbox"/> PATRIOTISM	4931	26	148	439	359	333	406	259	308	357	329	482	290	222	179	114	116	156	170	125	113
3	<input type="checkbox"/> COMMUNISM	4798				6	15	4	58	102	26	16	34	169	441	497	1451	940	292	279	321	147
4	<input type="checkbox"/> MECHANISM	4546		20	71	141	96	107	121	98	152	286	276	381	359	291	376	330	338	268	492	343
5	<input type="checkbox"/> SOCIALISM	3546				18	55	10	148	92	181	213	446	270	398	332	279	295	312	304	136	57
6	<input type="checkbox"/> ORGANISM	3426		3	2	69	36	82	175	230	179	321	289	374	273	256	343	191	214	137	120	132
7	<input type="checkbox"/> JOURNALISM	2633		3	1	57	17	34	87	60	99	108	151	196	131	201	177	183	207	292	283	346
8	<input type="checkbox"/> OPTIMISM	2517		1	12	5		6	22	45	49	104	165	237	225	188	238	219	226	278	195	302
9	<input type="checkbox"/> CAPITALISM	2513								3	7	12	65	135	271	247	218	208	259	453	436	199
10	<input type="checkbox"/> DESPOTISM	2265	23	84	204	293	388	287	199	120	119	90	86	103	55	48	47	19	27	42	23	8
11	<input type="checkbox"/> BAPTISM	2109		49	40	217	531	260	131	162	110	101	61	39	44	49	47	54	49	42	44	79
12	<input type="checkbox"/> HEROISM	2040	18	61	71	115	169	163	113	180	133	109	171	113	84	91	83	62	69	95	60	80
13	<input type="checkbox"/> REALISM	2018		5		1	13	24	49	123	120	123	112	198	147	140	237	165	116	152	158	135
14	<input type="checkbox"/> NATIONALISM	1847				1	1	3	4	44	24	15	99	203	172	232	196	264	141	182	170	96
15	<input type="checkbox"/> TERRORISM	1823			1	2	4	7	9	3	8	12	19	57	55	51	30	62	221	387	148	747

研究人员还可以在语料库中找到所有在某一时期比其它时期更常见的单词。例如，COHA 语料库显示了包含 **heart** (*heart* 字符串) 的单词在 19 世纪 (左图) 和 20 世纪晚期 (右图) 的频数对比，TIME 语料库显示了 **ess words** (以 *ess* 结尾的单词) 在 20 世纪 20 至 30 年代 (左图) 和 20 世纪 80 年代至 21 世纪 10 年代 (右图) 的频数对比；注意那些年代久远、带有女性词缀的单词，例如 *negress*、*authoress*、*sculptress*、*adventuress* 和 *poetess*。

SEC 1 (1820, 1830, 1840, 1850, 186...): 129,755,748 WORDS

WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1 HEART-STRINGS	188	0	1.4	0.0	144.9
2 NOBLE-HEARTED	132	1	1.0	0.0	108.5
3 HEARTH-STONE	135	0	1.0	0.0	104.0
4 HEART-BROKEN	346	3	2.7	0.0	94.8
5 HEART-SICK	114	0	0.9	0.0	87.9
6 HEARTSEASE	199	2	1.5	0.0	81.8
7 SINGLE-HEARTED	69	1	0.5	0.0	56.7
8 HEARTH-RUG	72	0	0.6	0.0	55.5
9 TRUE-HEARTED	199	3	1.5	0.0	54.5
10 HEART-ACHE	60	0	0.5	0.0	46.2
11 SIMPLE-HEARTED	160	3	1.2	0.0	43.8
12 HEART-BURNINGS	55	0	0.4	0.0	42.4

SEC 2 (1970, 1980, 1990, 2000): 106,640,094 WORDS

WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1 HEARTBEAT	664	3	6.2	0.0	269.3
2 HEARTLAND	273	0	2.6	0.0	256.0
3 WHOLEHEARTEDLY	152	1	1.4	0.0	184.9
4 HALFHEARTEDLY	68	1	0.6	0.0	82.7
5 MIND-AND-HEART	85	0	0.8	0.0	79.7
6 HEARTWARMING	60	1	0.6	0.0	73.0
7 HEART-STOPPING	56	0	0.5	0.0	52.5
8 OPEN-HEART	54	0	0.5	0.0	50.6
9 HEART-TO-HEART	48	0	0.5	0.0	45.0
10 HEARTTHROB	45	0	0.4	0.0	42.2
11 HEART-HEALTHY	39	0	0.4	0.0	36.6
12 HEART-ATTACK	38	0	0.4	0.0	35.6

SEC 1 (1930s, 1920s): 20,292,651 WORDS

WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1 CINEMACTRESS	139	0	6.8	0.0	685.0
2 NEGRESS	62	0	3.1	0.0	305.5
3 EYE-WITNESS	23	0	1.1	0.0	113.3
4 PROPRIETRESS	22	0	1.1	0.0	108.4
5 FESS	53	1	2.6	0.0	71.9
6 AUTHORESS	50	1	2.5	0.0	67.8
7 MARCHIONESS	50	1	2.5	0.0	67.8
8 SCULPTRESS	45	1	2.2	0.0	61.1
9 JEWESS	66	2	3.3	0.1	44.8
10 MARQUESS	181	7	8.9	0.3	35.1
11 SEASICKNESS	43	2	2.1	0.1	29.2
12 ADVENTURESS	21	1	1.0	0.0	28.5
13 POETESS	38	2	1.9	0.1	25.8
14 COUNTESS	517	32	25.5	1.2	21.9

SEC 2 (1980s, 1990s, 2000s): 27,534,890 WORDS

WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1 COMPETITIVENESS	116	0	4.2	0.0	421.3
2 SELF-AWARENESS	53	0	1.9	0.0	192.5
3 WEIRDNESS	43	0	1.6	0.0	156.2
4 WEIGHTLESSNESS	34	0	1.2	0.0	123.5
5 HIPNESS	32	0	1.2	0.0	116.2
6 AGRIBUSINESS	29	0	1.1	0.0	105.3
7 SEXINESS	27	0	1.0	0.0	98.1
8 PERMISSIVENESS	27	0	1.0	0.0	98.1
9 TOGETHERNESS	26	0	0.9	0.0	94.4
10 DEFENSIVENESS	23	0	0.8	0.0	83.5
11 FECKLESSNESS	22	0	0.8	0.0	79.9
12 DIVISIVENESS	21	0	0.8	0.0	76.3
13 OPENNESS	193	2	7.0	0.1	71.1
14 HOMELESSNESS	84	2	3.1	0.1	31.0

English-Corpora.org 还可用于研究语法的变迁。(自 COHA 语料库于 2010 年发布起) 的过去十年间，English-Corpora.org 广泛用于研究。例如，可用 COHA 查看过去 200 年间 *GET + V-ed* 这类短语 (例如 *get married*、*got painted*) 出现的频数，或查看 *END up V-ing* 这类短语 (例如 *ended up paying too much*) 的频数；可以看到，后者是近 100 年前才开始使用的。

SECTION	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	34125	18	98	368	374	561	622	877	926	860	1199	1516	1559	2048	2413	2610	2703	3001	3001	4365	5006
WORDS (M)	405	1.2	6.9	13.8	16.0	16.5	17.1	18.6	20.3	20.6	22.1	22.7	25.7	24.6	24.3	24.5	24.0	23.8	25.3	27.9	29.6
PER MIL	84.26	15.24	14.15	26.72	23.30	34.06	36.47	47.25	45.58	41.75	54.26	66.78	60.77	83.24	99.10	106.34	112.73	126.01	118.54	156.22	169.31
SEE ALL YEARS AT ONCE																					

END up V-ing

SECTION	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	1535	0	0	0	0	0	0	0	0	0	0	0	0	2	13	39	90	155	232	442	562
WORDS (M)	405	1.2	6.9	13.8	16.0	16.5	17.1	18.6	20.3	20.6	22.1	22.7	25.7	24.6	24.3	24.5	24.0	23.8	25.3	27.9	29.6
PER MIL	3.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.53	1.59	3.75	6.51	9.16	15.82	19.01
SEE ALL YEARS AT ONCE																					

研究人员也可通过搭配词研究语义的变化，因为搭配词的改变可能标志着一个单词含义或用法的变化。下图是过去 200 年各年代里 [gay](#) 的搭配词变化。在 19 世纪，gay 一词更多和“快乐、喜悦”联系在一起，到了 20 世纪后半叶，则向“性取向”靠拢。

HELP	<input type="checkbox"/>	CONTEXT	ALL	<input type="checkbox"/>	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
1	<input type="checkbox"/>	BRIGHT	172	1	5	8	10	14	13	23	12	14	12	4	12	12	8	11	7	4	2			
2	<input type="checkbox"/>	LESBIAN	153							1				1						1	5	67	78	
3	<input type="checkbox"/>	HAPPY	153		2	13	14	7	19	8	9	11	8	12	11	14	3	8	8	3	1	2		
4	<input type="checkbox"/>	FLOWERS	152		5	13	10	17	9	18	16	7	13	10	11	7	5	6	1	3		1		
5	<input type="checkbox"/>	LAUGH	137		2	7	5	15	13	12	14	7	12	14	8	11	2	4	7	4				
6	<input type="checkbox"/>	GRAVE	132		6	15	14	10	15	8	13	13	18	8	5	4	1	1						1
7	<input type="checkbox"/>	RIGHTS	129																	6	19	47	57	
8	<input type="checkbox"/>	COLORS	127		3	6	4	9	13	8	9	10	5	7	10	6	17	8	5	6	1			
9	<input type="checkbox"/>	GAY	100			4	2	2	10		8	10	6	2	4	16	4		6	6	2	10	8	
10	<input type="checkbox"/>	MARRIAGE	93				1		1	1					1					1		7	81	
11	<input type="checkbox"/>	LAUGHTER	88				5	5	6	6	8	3	6	15	9	11	3	2	3	2	3	1		
12	<input type="checkbox"/>	GALLANT	87	1	7	11	12	4	9	7	9	6	6	1	9	3	1	1						
13	<input type="checkbox"/>	BRILLIANT	75		3	8	7	10	8	7	5	3	5	4	3	3	5	4						
14	<input type="checkbox"/>	VOICES	71			1	2	7	8	4	10	2	5	8	1	5	10	4	3		1			
15	<input type="checkbox"/>	CHEERFUL	65		2	6	5	6	5	5	7	2	5	6	6		4	2		2	2			

搭配词也可以标志我们对某一话题讨论内容的变化。例如，19 世纪的文本中 [women](#) 的搭配词（左图）带有强烈的性别歧视意味：从道德角度对女性评价（*noble*、*true*、*pure*、*cultivated*、*refined*、*wretched*）；通常认为女性处于弱势（*unfortunate*、*abandoned*、*helpless*）；给聪明或独立的女性打上“不寻常”的标签（*strong-minded*、*clever*）。

SEC 1 (1820, 1830, 1840, 1850, 186...): 174,553,979 WORDS

SEC 2 (1970, 1980, 1990, 2000): 106,640,094 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	STRONG-MINDED WOMEN	24	1	0.1	0.0	14.7	1	PREGNANT WOMEN	233	5	2.2	0.0	76.3
2	CLEVER WOMEN	24	0	0.1	0.0	13.7	2	BATTERED WOMEN	70	0	0.7	0.0	65.6
3	NOBLE WOMEN	36	2	0.2	0.0	11.0	3	AFRICAN-AMERICAN WOMEN	61	0	0.6	0.0	57.2
4	TRUE WOMEN	18	1	0.1	0.0	11.0	4	DIVORCED WOMEN	25	1	0.2	0.0	40.9
5	UNFORTUNATE WOMEN	17	1	0.1	0.0	10.4	5	MIDDLE-CLASS WOMEN	23	1	0.2	0.0	37.6
6	WRETCHED WOMEN	18	0	0.1	0.0	10.3	6	MUSLIM WOMEN	23	1	0.2	0.0	37.6
7	ABANDONED WOMEN	18	0	0.1	0.0	10.3	7	NATIONAL WOMEN	68	3	0.6	0.0	37.1
8	HELPLESS WOMEN	66	4	0.4	0.0	10.1	8	BLACK WOMEN	487	22	4.6	0.1	36.2
9	VERY WOMEN	15	1	0.1	0.0	9.2	9	MENOPAUSAL WOMEN	22	1	0.2	0.0	36.0
10	TURKISH WOMEN	15	1	0.1	0.0	9.2	10	SOVIET WOMEN	32	0	0.3	0.0	30.0
11	ELDER WOMEN	15	0	0.1	0.0	8.6	11	ADULT WOMEN	18	1	0.2	0.0	29.5
12	DEFENCELESS WOMEN	15	0	0.1	0.0	8.6	12	IMMIGRANT WOMEN	15	1	0.1	0.0	24.6
13	AGED WOMEN	28	2	0.2	0.0	8.6	13	AFGHAN WOMEN	26	0	0.2	0.0	24.4
14	FAIR WOMEN	69	5	0.4	0.0	8.4	14	MISSING WOMEN	14	1	0.1	0.0	22.9
15	PURE WOMEN	14	0	0.1	0.0	8.0	15	SUCCESSFUL WOMEN	14	1	0.1	0.0	22.9
16	HANDSOME WOMEN	37	3	0.2	0.0	7.5	16	GOOD-LOOKING WOMEN	13	1	0.1	0.0	21.3
17	CULTIVATED WOMEN	13	0	0.1	0.0	7.4	17	MATURE WOMEN	13	1	0.1	0.0	21.3
18	REFINED WOMEN	12	0	0.1	0.0	6.9	18	LOCAL WOMEN	22	0	0.2	0.0	20.6

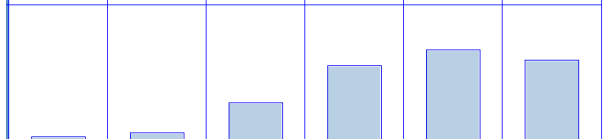
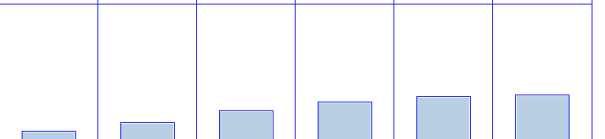
除 English-Corpora.org 的语料库以外，没有其它任何历时语料库有足够丰富的库容和强大的架构，去完成以上两例搭配词的搜索。而且，在收词 4 亿条的 COHA 或是 English-Corpora.org 的其它历时语料库里，要实现以上复杂的搜索并提供大量有用的数据，仅需 1 至 2 秒。

近期变化 [\(返回开头\)](#)


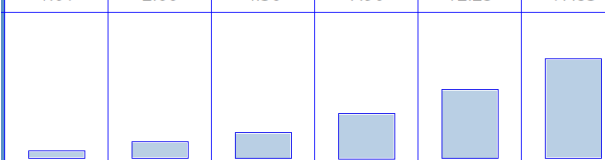
EEBO、COHA、US Supreme Court，以及 Hansard（英国议会）这几个语料库聚焦数百年前或是近两百年左右英语的变化。但 English-Corpora.org 系列语料库还有个特点，即能让研究人员查看英语近些年的变化。电影语料库（涵盖 1930-2010 年）和电视语料库（涵盖 1950-2010 年）是**全球仅有的两个大型语料库**，**聚焦近年非正式英语口语的变化**。例如，它们能显示 1930—1969 年（左栏）和 1990—2018 年（右栏）更常见的单词（包括近几十年电影中出现较多的秽语）。

	More common 1930-1969 (movies)	More common 1990-2018 (movies)
ADJ	swell, splendid, sore, fond, delighted, dreadful, darn, phony, blasted, satisfactory, snappy, darned, apt, no-good, cockeyed, screwy, disgraceful, crummy, beastly, frightful, double-crossing, phoney, bashful, confounded, shrewd, soapy, daffy	f--king, okay, cool, weird, damn, g--d---, huge, awesome, pregnant, super, sexy, scary, unbelievable, sexual, boring, pathetic, gross, massive, nuclear, creepy, global, creative, magical, intense, ultimate, sh-tty, homeless, random, corporate, pissed
NOUN	darling, fellow, pardon, dough, wagon, headquarters, chap, cigar, railroad, brandy, telegram, corporal, crook, hunch, regiment, squadron, handkerchief, shilling, cinch, butler, skipper, chauffeur, plenty, tailor, sonny, mink, nuisance, mammy, waltz, newspaperman	sh-t, hell, mom, f--k, a-s, b-tch, dude, sex, drug, a--h---, tv, bullsh-t, m-f-r, b-st-rd, girlfriend, relationship, d-ck, computer, video, tape, crap, bro, p-ssy, n-g--, grunt, role, bike, chick, cancer, butt
VERB	shall, suppose, pardon, phone, spoil, frighten, telephone, permit, object, congratulate, oblige, dine, notify, faint, quarrel, acquaint, delight, amuse, intrude, dislike, slug, scram, furnish, sock, darn, consent, tangle, fuss, peddle, double-cross	f--k, suck, screw, p-ss, focus, freak, date, r-pe, pee, film, score, b-tch, sh-t, chill, define, stress, evolve, f-rt, activate, surf, tape, participate, process, monitor, target, manipulate, trigger, puke, initiate, generate

在 COCA 中如何查看英语基于不同体裁的变体，已经在前面小节做了介绍。不过，从 1990 到 2019 年，该语料库里各类体裁的比例每年几乎没有变化，因此这个拥有十亿库容的语料库也可以用于查看**过去 30 年英语的变化**（COCA 是世界上唯一能实现这类搜索的语料库）。例如，用户可按五年为一个时间段查看单词和短语的频数（也可查看每年的频数）。下图展示了以五年为单位，搜索 *old-school* 和 *freak out*，能看到它们使用频数的增加（现在是 25-30 年前的四倍多）。

old-school						freak out					
1990-94	1995-99	2000-04	2005-09	2010-14	2015-19	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
26	48	209	397	483	426	246	479	788	998	1121	1158
139.1	147.8	146.6	144.9	145.3	144.7	139.1	147.8	146.6	144.9	145.3	144.7
0.19	0.32	1.43	2.74	3.33	2.94	1.77	3.24	5.38	6.89	7.72	8.00
											

研究人员也可研究英语近年来的句法变化，例如 END up V-ing 结构（如 *we ended up leaving at 9 AM instead*）或是 like 结构（如 *and I was like, I guess they can come*）使用的增加。

END up V-ing						“like construction”					
1990-94	1995-99	2000-04	2005-09	2010-14	2015-19	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
1826	2340	2489	2849	2949	3292	140	393	639	1145	1780	2581
139.1	147.8	146.6	144.9	145.3	144.7	139.1	147.8	146.6	144.9	145.3	144.7
13.13	15.83	16.98	19.66	20.30	22.74	1.01	2.66	4.36	7.90	12.25	17.83
											

前文以单词 *gay* 为例，演示了在 COHA 语料库中如何通过查看搭配词来发现其语义的变化。在 COCA 中，也可进行同样的搜索，查看单词在过去 30 年里语义和用法的改变。以下分别是 [web 的搭配词](#) 的搜索结果（注意：从 20 世纪 90 年代初开始，与万维网相关的单词开始增加（右上图）），以及 [green 的名词搭配词](#) 的结果，显示 21 世纪 10 年代（右下图）*green* 有了更新的含义，即“环保”。

SEC 1 (1990-1994): 139,059,192 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	SPIDER	142	219	1.0	0.5	2.0
2	LIFE	39	86	0.3	0.2	1.4
3	RELATIONSHIPS	20	45	0.1	0.1	1.4
4	FOOD	25	187	0.2	0.4	0.4

SEC 2 (2005-2009, 2010-2014, 2015-...): 434,948,338 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	SITE	8830	2	20.3	0.0	1,411.5
2	SITES	2176	2	5.0	0.0	347.8
3	PAGE	633	2	1.5	0.0	101.2
4	PAGES	414	0	1.0	0.0	95.2
5	SEARCH	366	0	0.8	0.0	84.1
6	E-MAIL	356	0	0.8	0.0	81.8
7	BROWSER	301	0	0.7	0.0	69.2
8	VIDEO	194	1	0.4	0.0	62.0
9	COMPANY	191	1	0.4	0.0	61.1
10	ADDRESS	186	1	0.4	0.0	59.5
11	RESOURCES	167	1	0.4	0.0	53.4

SEC 1 (1995-1999, 1990-1994): 286,833,557 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	GREEN PEPPER	215	35	0.7	0.1	6.2
2	GREEN CROSS	52	9	0.2	0.0	5.8
3	GREEN VEGETABLES	73	32	0.3	0.1	2.3
4	GREEN PEPPERS	94	47	0.3	0.2	2.0
5	GREEN MAN	58	31	0.2	0.1	1.9
6	GREEN ACRES	52	31	0.2	0.1	1.7
7	GREEN BELL	154	92	0.5	0.3	1.7
8	GREEN PLANTS	61	37	0.2	0.1	1.7
9	GREEN GLASS	61	37	0.2	0.1	1.7
10	GREEN WATER	120	74	0.4	0.3	1.6
11	GREEN BERETS	69	46	0.2	0.2	1.5
12	GREEN MONSTER	57	39	0.2	0.1	1.5

SEC 2 (2010-2014, 2015-2019): 290,003,115 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	GREEN GAZETTE	96	0	0.3	0.0	33.1
2	GREEN JOBS	87	0	0.3	0.0	30.0
3	GREEN PRACTICE	60	2	0.2	0.0	29.7
4	GREEN ENERGY	170	7	0.6	0.0	24.0
5	GREEN ARROW	192	8	0.7	0.0	23.7
6	GREEN BUILDING	130	18	0.4	0.1	7.1
7	GREEN SCREEN	85	12	0.3	0.0	7.0
8	GREEN ZONE	118	21	0.4	0.1	5.6
9	GREEN LANTERN	96	21	0.3	0.1	4.5
10	GREEN SPACES	97	23	0.3	0.1	4.2
11	GREEN BANK	84	21	0.3	0.1	4.0
12	GREEN MOVEMENT	93	25	0.3	0.1	3.7

NOW 语料库的独到之处就是能显示英语新近变化。该语料库共收集了从 2010 年至今（确切地说，2020 年 11 月本指南完稿前一天）约 115 亿单词。每一天，该语料库会新增 600 万至 1000 万单词，也就是每

月新增 2 至 2.5 亿单词。用户可查看单词和短语每半年的频数（甚至每十天的频数）。例如，下面截图显示 [fake news](#) 在 2016 年下半年（即红框中的 2016-2）使用激增，可更具体地看到，该短语是在 11 月上旬到 11 月中旬之间出现激增，紧随当年 11 月 8 日的美国大选。



NOW 语料库还可显示过去十年短语的变化，例如对比 2010-2013 年（左图），[data + NOUN](#) 短语在 2018-2020 年（右图；如 *data ethics*、*data scandal*）使用更频繁。

SEC 1 (2010-1, 2010-2, 2011-1, 2011-2, 2012-1, 2012-2, 2013-1, 2013-2, 2014-1, 2014-2, 2015-1, 2015-2, 2016-1, 2016-2, 2017-1, 2017-2, 2018-1, 2018-2, 2019-1, 2019-2, 2020-1, 2020-2)						SEC 2 (2018-1, 2018-2, 2019-1, 2019-2, 2020-1, 2020-2)					
WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1 DATA BYTE	78	1	0.0	0.0	256.8	1 DATA ETHICS	474	1	0.1	0.0	144.0
2 DATA APPLIANCE	38	1	0.0	0.0	125.1	2 DATA SCANDAL	852	2	0.1	0.0	129.4
3 DATA DISASTERS	70	7	0.0	0.0	32.9	3 DATA TRIANGULATION	285	1	0.0	0.0	86.6
4 DATA SENSE	37	4	0.0	0.0	30.5	4 DATA SAVER	216	1	0.0	0.0	65.6
5 DATA FILES	9945	1332	5.7	0.2	24.6	5 DATA STORYTELLING	165	1	0.0	0.0	50.1
6 DATA PROMOTIONS	50	10	0.0	0.0	16.5	6 DATA SCANDALS	148	1	0.0	0.0	44.9
7 DATA FEDERATION	31	9	0.0	0.0	11.3	7 DATA BIAS	116	1	0.0	0.0	35.2
8 DATA DEVICES	55	24	0.0	0.0	7.5	8 DATA BENEFIT	110	1	0.0	0.0	33.4
9 DATA STREAM	587	258	0.3	0.0	7.5	9 DATA LITERACY	767	7	0.1	0.0	33.3
10 DATA MARTS	36	21	0.0	0.0	5.6	10 DATA PREFERENCES	87	1	0.0	0.0	26.4
11 DATA DEVICE	32	20	0.0	0.0	5.3	11 DATA LOCALISATION	1475	0	0.3	0.0	25.6

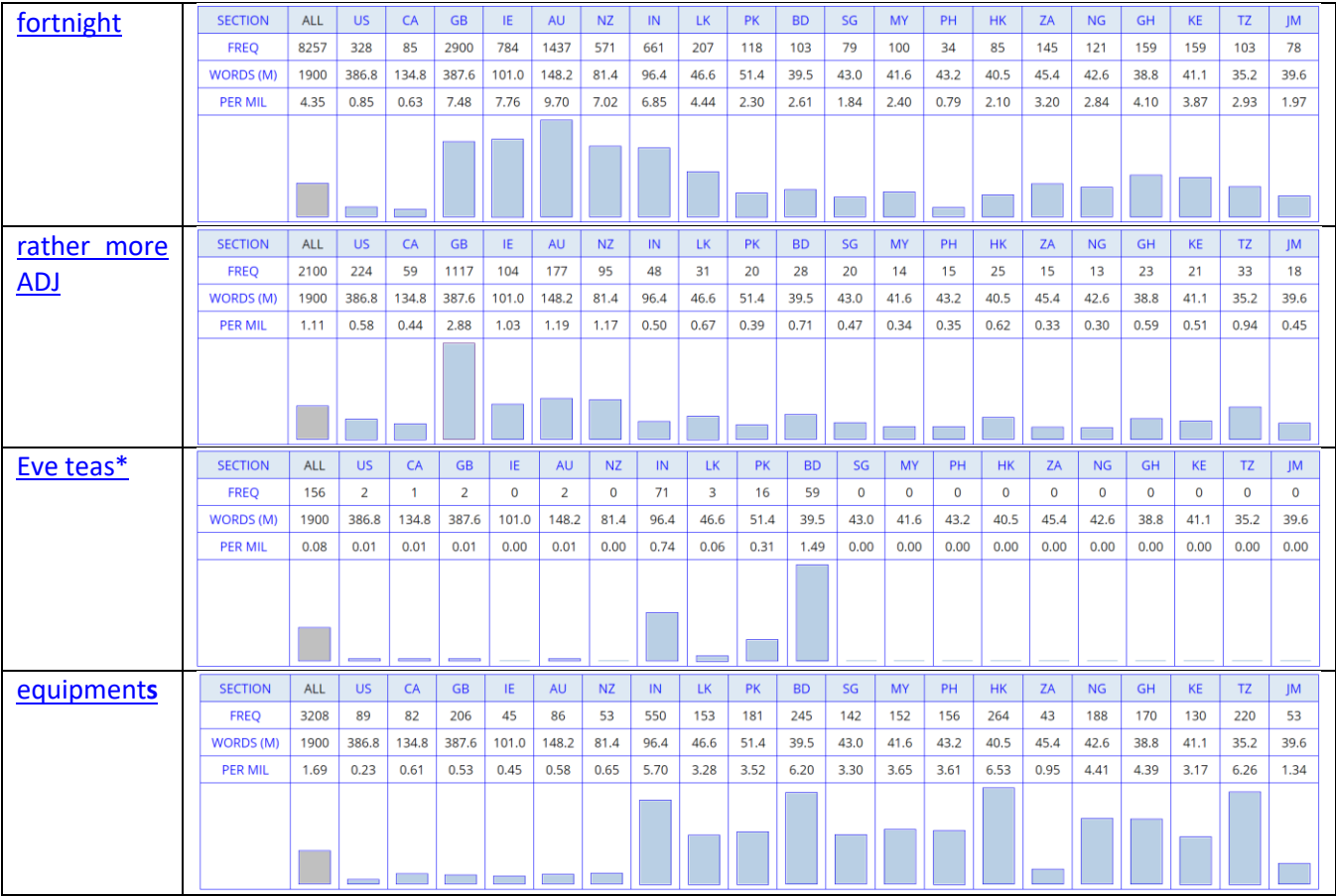
Coronavirus 语料库 是 NOW 语料库的子集，包含自 2020 年起关于新型冠状病毒的文章。截至 2020 年末，Coronavirus 语料库约包含 7 亿单词，每月还会新增约 6000 万至 7000 万单词。它能显示自 2020 年 1 月起，单词和短语每十天的频数。例如，我们可以看到 [flatten the curve](#) 的频数在 2020 年 3 月中旬达到最高，6 月起趋于平缓。

SECTION	ALL	20-01-01	20-02-01	20-02-11	20-02-21	20-03-01	20-03-11	20-03-21	20-04-01	20-04-11	20-04-21	20-05-01	20-05-11	20-05-21	20-06-01	20-06-11	20-06-21
FREQ	11701	0	0	0	0	35	533	2117	1566	1264	1029	773	647	531	350	333	352
WORDS (M)	237	7.3	4.8	4.0	5.7	17.6	26.8	55.5	38.4	35.8	33.8	31.3	30.5	36.1	29.3	27.1	26.8
PER MIL	49.24	0.00	0.00	0.00	0.00	1.98	19.88	38.12	40.76	35.35	30.44	24.73	21.22	14.72	11.94	12.28	13.12

方言变体 (返回开头)

GloWbE 语料库收入 20 个不同英语国家大约 20 亿词的语料。研究人员可以通过该语料库查看不同方言的差异，而其它任何一个语料库都达不到同样的效果。自 GloWbE 2013 年公开以来，研究人员已发表了大量基于该语料库的文章。

最基本的功能是，研究人员可以查看某个单词或词组在这 20 个国家的使用频数，例如 *fortnight*（注意，这个单词在美国和加拿大英语，以及基于美语的菲律宾英语中基本没有使用），*rather more ADJ*（显然在英国英语中使用得最频繁），*Eve teas**（意思是“性骚扰”，基本只在南亚使用），以及 *equipments*（复数形式）在大多数国家都较多使用，除六个“核心”国家外（指美国、加拿大、英国、爱尔兰、澳大利亚和新西兰）。



用户也可以查询匹配特定字符串的各种单词在全部 20 个国家的使用频数。例如，下图所示是使用最多的 [*ism words \(以 ism 结尾的单词\)](#)。

HELP	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
1	TOURISM	66231	2862	3177	7376	3290	4237	3871	3564	3718	922	1706	2138	2451	2314	2950	2637	1094	2838	3746	6370	4970
2	CRITICISM	62753	14465	3646	15809	3165	4984	2298	3018	1841	2200	1148	811	1022	816	1125	1451	1316	1037	968	721	912
3	MECHANISM	44354	8851	2554	8022	2293	3576	1793	3275	1737	1107	1178	886	920	705	1636	1065	760	830	1345	1067	754
4	TERRORISM	42215	8783	1912	6845	732	2102	882	2941	5427	5530	1570	317	472	318	417	397	1279	463	1024	544	260
5	JOURNALISM	41483	10282	2879	10441	1591	3954	1090	1695	998	746	929	522	336	613	648	842	786	908	896	865	462
6	CAPITALISM	37344	9466	2269	10261	1944	2835	1551	1358	683	603	874	461	220	368	875	850	517	394	372	819	624
7	RACISM	36556	11535	1896	8545	1860	2988	1052	797	1082	579	332	503	832	199	327	1185	586	676	508	368	706
8	BUDDHISM	21816	1830	310	1437	351	757	390	1791	9064	324	829	846	1205	314	1955	76	66	70	58	87	56
9	AUTISM	20350	7250	1514	5285	1590	2211	264	715	76	58	274	73	98	160	106	66	41	77	37	72	383
10	SOCIALISM	19851	6427	792	4292	1020	1732	734	746	292	284	536	192	114	225	534	413	174	202	156	690	296
11	OPTIMISM	15144	2950	1251	3767	767	990	533	678	265	375	324	347	244	328	303	297	364	379	483	242	257
12	NATIONALISM	14409	1523	880	3053	1022	851	270	1033	1474	887	773	143	186	287	310	368	347	277	213	230	282
13	COMMUNISM	14216	4466	630	3286	632	1249	401	504	190	321	377	204	227	235	330	395	161	118	132	208	150
14	BAPTISM	12386	2697	1506	1315	967	918	814	193	179	83	795	130	89	696	253	285	224	572	166	302	202
15	FEMINISM	12235	4159	887	2932	557	1491	484	249	124	126	96	61	92	78	54	152	257	139	166	84	47

电视语料库和电影语料库也能提供方言差异方面的有用信息，因为它们汇集了来自六个“核心”国家的非常口头的英语语料，库容达 5.75 亿词。例如，下表所示是对应各词性，美国英语和英国英语中更常见的单词。当然，这两个语料库也能比较这六个国家方言间任何其它方面的差异，包括构词、句法、语义和用法（通过“搭配词”体现）。

	American	British
ADJ	okay, crazy, damn, awesome, cute, dumb, federal, goddamn, gross, lame, adorable, lousy, crappy, sloppy, phony, downtown, cozy, busted, darn, cranky, high-end, one-time, high-school, canned, cellular, big-time, African-American, goofy, off-limits, old-school, sassy, condescending, puffy, big-a--, sketchy, wordy, charmed, disoriented, kick-a--, bitchy, narcissistic, crummy, self-centered, curt, trashy, whimsical, dorky, scrappy	daft, posh, dodgy, knackered, ruddy, barmy, sodding, poxy, dozy, soppy, mucky, disused, chuffed, tinned, whirly, manky, disorientated, pish, fiddly
NOUN	guy, mom, honey, dude, cop, agent, a--, movie, buddy, apartment, truck, chef, buck, dollar, sweetie, mommy, attorney, mayor, butt, cookie, grandma, a--h---, candy, grade, parking, senator, couch, vacation, closet, homicide, garbage, jerk, baseball, grandpa, elevator, trash, math, thanksgiving, shooter, roommate, bud, assignment, prom, tech, mall, dessert, heck, bout, zombie, soda, motel, halloween, therapist, basketball, counselor, lawsuit, diaper, congressman, chili,	mum, bloke, a-se, quid, rubbish, b-l-ck, solicitor, railway, vicar, telly, guv, grandad, petrol, ladyship, mammy, shilling, maths, lorry, a---h---, advert, motorway, tosser, tenner, pence, nutter, punter, gearbox, footballer, windscreen, pensioner, barman, pram, tuppence, prat, flatmate, lodger, roundabout, vicarage, workhouse, pillock, sixpence
VERB	guess, figure, kid, damn, date, quit, hire, freak, yell, bust, file, hook, testify, pee, coach, assign, schedule, graduate, violate, practice, dial, jerk, sniffle, participate, brag, party, merge, poop, hustle, reschedule	reckon, fancy, shag, sod, flog, w-nk, queue, burgle, snigger, snog, plod, splutter, clamber

大量研究也使用了 GloWbE 语料库考察不同方言的句法差异，这里举两个例子予以说明。[“like 结构”](#)（如 *and I'm like, no way can they do it*）在美国英语中使用得最频繁，但也出现在其它“核心”国家，如加拿大、英国、爱尔兰、澳大利亚和新西兰（尽管频数从前往后越来越低）。下面右图是 [try and VERB 结构](#)（如 *I'm gonna try and talk to her*，注意对比 *try to talk*）在六国的使用频数，该结构在美国英语和加拿大英语中常被认为是“错误的”（源于这两个国家 50-100 年前约定俗成的语法），而在其它国家（不把这一语法规则奉为金科玉律的地方），该结构就常见多了。

“like” construction									try and VERB								
SECTION	ALL	US	CA	GB	IE	AU	NZ	IN	SECTION	ALL	US	CA	GB	IE	AU	NZ	
FREQ	2620	897	264	599	95	163	63	51	FREQ	65002	10321	3678	20649	4245	7201	3653	
WORDS (M)	1900	386.8	134.8	387.6	101.0	148.2	81.4	96.4	WORDS (M)	1900	386.8	134.8	387.6	101.0	148.2	81.4	
PER MIL	1.38	2.32	1.96	1.55	0.94	1.10	0.77	0.53	PER MIL	34.21	26.68	27.29	53.27	42.02	48.59	44.88	

基于 GloWbE 语料库的庞大库容（接近 20 亿词），我们也可以用搭配词来考查不同方言在语义和用法方面的差异。例如，下图所示是 [scheme 的搭配词](#)，其中 *alleged*, *evil*, *fraudulent*, *nefarious* 等搭配说明该词在美国英语中的消极色彩比在英国英语中更浓烈。

SEC 1 (United States): 386,809,355 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	BLOCKING	42	1	0.1	0.0	42.1
2	URI	80	6	0.2	0.0	13.4
3	OFFENSIVE	61	6	0.2	0.0	10.2
4	CONSTITUTIONAL	16	2	0.0	0.0	8.0
5	DEFENSIVE	89	13	0.2	0.0	6.9
6	SOCIALIST	20	3	0.1	0.0	6.7
7	ALLEGED	26	5	0.1	0.0	5.2
8	EVIL	48	10	0.1	0.0	4.8
9	LEGISLATIVE	15	4	0.0	0.0	3.8
10	FRAUDULENT	62	18	0.2	0.0	3.5
11	NEFARIOUS	27	9	0.1	0.0	3.0
12	PONZI	617	255	1.6	0.7	2.4

SEC 2 (Great Britain): 387,615,074 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	APPROVED	92	1	0.2	0.0	91.8
2	OCCUPATIONAL	88	1	0.2	0.0	87.8
3	MENTORING	53	1	0.1	0.0	52.9
4	FLAT	36	1	0.1	0.0	35.9
5	ELIGIBLE	31	1	0.1	0.0	30.9
6	OVERSEAS	31	1	0.1	0.0	30.9
7	DEFINED	127	5	0.3	0.0	25.3
8	GENEROUS	50	2	0.1	0.0	24.9
9	LABOUR	25	1	0.1	0.0	24.9
10	TAX-AVOIDANCE	25	1	0.1	0.0	24.9
11	SCOTTISH	24	1	0.1	0.0	24.0
12	INNOVATIVE	70	3	0.2	0.0	23.3

对于讲不同方言的人们如何谈论不同话题，我们也可以用搭配词来对比，可揭示文化和社会层面一些有意思的差异。例如，亚洲和非洲方言中 [wife 的搭配词](#)（左图）包括 *existing*, *temporary*, *permanent* 等，都与这些国家的文化习俗有关。像 *chaste*, *obedient*, *good*, *virtuous* 等其它搭配词也传达了这些国家重要的文化习俗和社会规范。可以看到，一次仅用 2、3 秒的简单搜索——如果用对话料库的话——就能呈现出不同国家文化的一些有意思的差异，不止语言学家，社会学家也会对此感兴趣。

SEC 1 (India, Sri Lanka, Pakistan,...): 644,753,594 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	EXISTING WIFE	25	1	0.0	0.0	48.1
2	CHASTE WIFE	21	1	0.0	0.0	40.4
3	PAKISTANI WIFE	23	3	0.0	0.0	14.7
4	SENIOR WIFE	21	3	0.0	0.0	13.5
5	TEMPORARY WIFE	27	4	0.0	0.0	13.0
6	OBEDIENT WIVES	23	6	0.0	0.0	7.4
7	PERMANENT WIFE	45	0	0.1	0.0	7.0
8	MUSLIM WIFE	94	26	0.1	0.0	7.0
9	AFRICAN WIFE	20	7	0.0	0.0	5.5
10	DIVORCED WIFE	41	15	0.1	0.0	5.3
11	LEGAL WIFE	72	27	0.1	0.0	5.1
12	WEDDED WIFE	54	22	0.1	0.0	4.7
13	OTHER WIFE	109	48	0.2	0.0	4.4
14	POTENTIAL WIFE	36	16	0.1	0.0	4.3
15	BEAUTIFUL WIVES	22	10	0.0	0.0	4.2
16	MARRIED WIFE	40	20	0.1	0.0	3.8
17	GOOD WIVES	51	26	0.1	0.0	3.8
18	VIRTUOUS WIFE	25	13	0.0	0.0	3.7

SEC 2 (United States, Canada, Grea...): 1,239,817,686 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	PLURAL WIVES	35	1	0.0	0.0	18.2
2	DESERTED WIFE	68	3	0.1	0.0	11.8
3	GLAMOROUS WIFE	20	1	0.0	0.0	10.4
4	MILITARY WIVES	172	11	0.1	0.0	8.1
5	MILITARY WIFE	111	14	0.1	0.0	4.1
6	DESERTED WIVES	22	3	0.0	0.0	3.8
7	PLURAL WIFE	20	3	0.0	0.0	3.5
8	DYING WIFE	31	6	0.0	0.0	2.7
9	ILL WIFE	29	6	0.0	0.0	2.5
10	DISABLED WIFE	23	5	0.0	0.0	2.4
11	MERRY WIVES	50	11	0.0	0.0	2.4
12	POLITICAL WIVES	29	0	0.0	0.0	2.3
13	THEN WIFE	89	20	0.1	0.0	2.3
14	MISSING WIFE	26	6	0.0	0.0	2.3
15	AMAZING WIFE	62	15	0.1	0.0	2.1
16	HOT WIFE	44	11	0.0	0.0	2.1
17	AWESOME WIFE	23	6	0.0	0.0	2.0
18	IRISH WIFE	24	0	0.0	0.0	1.9

虚拟语料库 (返回开头)

从前面小节中可以看到，本语料库集划分了多个版块——例如体裁、年代和国家，供研究人员使用。用户也可以快速轻松地在语料库中创建一个文本集合，然后搜索这个“虚拟语料库”，就像是在用一个现成的语料库一样。比方说，用户可以专注于某一特定话题的文本（如生物、投资、核能、篮球或者哈利波特）、某一特定作者或文本来源（如《纽约时报》、《天文学杂志》）、某一特定体裁（如电视语料库中的真人秀、COCA 语料库或 BNC 语料库中的财经文章等）、某一特定时间段，或者任意组合以上选项。

下图具体展示了研究人员在电视语料库（左）和 NOW 语料库（右图上方）中创建虚拟语料库的页面。English-Corpora.org 的 17 个语料库每个都可以用这样类似的页面创建虚拟库。用户也可以只基于单词和词组快捷地创建虚拟语料库（右图下方）。

TV Corpus

SORT

Criteria

Values

Series title

Can use wildcards, e.g. *Star Trek*

Year

1950

-

2017

Genre

Drama (41644)

Comedy (31026)

Crime (17068)

Action (14314)

Adventure (1309)

Fantasy (6097)

Family (5805)

Sci-Fi (4481)

Documentary (1837)

History (1606)

Game-Show (1224)

Music (1183)

War (1153)

Biography (456)

Talk-Show (268)

News (230)

Musical (187)

Country

USA

Canada

UK

Ireland

Australia

New Zealand

Primary

Anywhere

TV rating

TV-14 (18692)

TV-PG (14204)

TV-MA (7061)

TV-G (1767)

TV-Y7 (1720)

ATP (157)

13 (121)

M (80)

16 (60)

15 (58)

6 (56)

N/A (132)

APPROVED (64)

IMDB rating

Low

High

(Min # votes)

1

Plot

(words in episode plot)

Word in text

(single word only)

Submit

Reset

Web domain

Guardian

Find sources

(can use substring, e.g. Times, Houston)

Article title

refugees

Country

United States

Canada

Great Britain

Ireland

Dates

09/01/2018

to

10/31/2018

Words in text

texts (max)

1000

Submit

Reset

List

Chart

Word

Browse

+

INVESTMENT

[POS]

Find matching strings

Reset

Sections

Texts/Virtual

Sort/Li

FIND TEXTS

MY CORPORA

Create corpus

接下来，语料库会显示出自认为最符合检索内容的文本，以供用户选择。用户可以增加或删除虚拟语料库中的文本，也可以在不同虚拟语料库间复制或移动文本。

HELP	<input type="checkbox"/>	100	TEXT	# WORDS	# HITS ↓	RELEVANCE ↓	PER MILLION WORDS
1	<input checked="" type="checkbox"/>		ACAD: THE JOURNAL OF CORPORATION LAW: INVESTORS' PARADOX	25682	322	12,538.0	<div></div>
2	<input checked="" type="checkbox"/>		ACAD: ENERGYJOURNAL: MARKET BARRIERS TO ENERG...	8693	181	20,821.4	<div></div>
3	<input checked="" type="checkbox"/>		BLOG: MPETTIS.COM: HOW TO BE A CHINA BULL	16037	133	8,293.3	<div></div>
4	<input checked="" type="checkbox"/>		ACAD: INTLAFAIRS: TRADE-RELATED INVESTMENT...	9199	132	14,349.4	<div></div>
5	<input checked="" type="checkbox"/>		ACAD: BYU LAW REV: TRUSTS NO MORE: RETHINKI...	23103	129	5,583.7	<div></div>
6	<input checked="" type="checkbox"/>		ACAD: CURRENT POLITICS AND ECONOMICS OF SOUTH, SOUTHE...: UZBEKISTAN: INVESTMENT C...	11398	108	9,475.3	<div></div>

用户可以看到所有自建的虚拟语料库，并且能按自定义类别进行归类（如科学、财经和体育）。

HELP		↑	↓	LIST NAME ↑	# ARTICLES ↑	# WORDS ↑	FIND KEYWORDS <input checked="" type="radio"/> SPECIFIC <input type="radio"/> FREQ
1	<div></div>	<div></div>	Sp	BASEBALL	100	413,279	NOUN VERB ADJ ADV N+N ADJ+N
2	<div></div>	<div></div>		BASKETBALL	100	257,867	NOUN VERB ADJ ADV N+N ADJ+N
3	<div></div>	<div></div>	Bi	BIOLOGY	100	142,355	NOUN VERB ADJ ADV N+N ADJ+N
4	<div></div>	<div></div>	Sc	BRAIN	100	132,983	NOUN VERB ADJ ADV N+N ADJ+N
5	<div></div>	<div></div>		BUDDHISM	100	228,673	NOUN VERB ADJ ADV N+N ADJ+N

19

也许最重要的是，用户可以查看自建虚拟语料库中的**主题词表**，还可以调整主题词对于该虚拟库的特有程度。以下是基于维基百科语料库自建的[生物]虚拟语料库中的主题词。

BIOLOGY2020 [155,354 WORDS, 100 TEXTS] **NOUN** VERB ADJ ADV N+N ADJ+N [ALL CORPORA] SAVE LIST

HELP	WORD (CLICK FOR CONTEXT)	FREQ	# TEXTS	<div><div><div><div>SPECIFIC</div></div><div>FREQ</div><div>30</div><div>10</div><div>TEXTS</div></div></div>	ALL WIKIPEDIA	EXPECTED
1	EUKARYOTE	34	11	1,984.7	204	0.0
2	MICROORGANISM	65	20	1,554.3	498	0.0
3	ORGANELLE	35	12	936.6	445	0.0
4	ORGANISM	378	60	365.2	12,327	1.0
5	MRNA	64	10	191.0	3,991	0.3
6	NEURON	42	13	122.1	4,097	0.3
7	BIOLOGIST	86	26	109.2	9,379	0.8
8	BIOLOGY	425	53	101.6	49,803	4.2
9	MOLECULE	114	31	86.2	15,753	1.3
10	ECOSYSTEM	64	13	81.9	9,303	0.8
11	ALGAE	64	25	80.6	9,459	0.8
12	MEMBRANE	148	18	79.7	22,106	1.9

用户单击某一主题词时，就能看到它在该虚拟语料库中索引行：

CLICK FOR MORE CONTEXT

☐ [?]

SAVE LIST

CHOOSE LIST

CREATE NEW LIST

[?]

SHOW DUPLICATES

1	Biological determinism	A B C	gender category, however, humans decide whether a person with XXY chromosomes or XY chromosomes and androgen insensitivity will count as intersex. # Soci
2	Cell (biology)	A B C	DNA molecules called chromosomes, including 22 homologous chromosome pairs and a pair of sex chromosomes . The mitochondrial genome is a circular DNA m
3	Polymorphism (biology)	A B C	a restricted food supply heterozygotes had a distinct advantage. 3. Different proportions of chromosome morphs were found in different areas. There is, for exan
4	Cell (biology)	A B C	. Prokaryotic genetic material is organized in a simple circular DNA molecule (the bacterial chromosome) in the nucleoid region of the cytoplasm. Eukaryotic genet
5	Hybrid (biology)	A B C	abnormalities #a numerical hybrid results from the fusion of gametes having different haploid numbers of chromosomes #a permanent hybrid is a situation whei
6	Synthetic biology	A B C	present new orthogonal functions in living cells. Genetic engineering includes approaches to construct synthetic chromosomes for whole or minimal organisms. B
7	Hybrid (biology)	A B C	(where the two times two comes about from two rounds of meiosis with two chromosomes); however, this probability declines markedly with chromosome numb
8	Hybrid (biology)	A B C	allopolyploidy occurs when two different species mate and produce polyploid hybrids. Usually the typical chromosome number is doubled, and the four sets of ch
9	Hybrid (biology)	A B C	their origins in polyploidy. Autopolyploidy results from the sudden multiplication in the number of chromosomes in typical normal populations caused by unsucc
10	Developmental biology	A B C	result in birth defects or miscarriage. Often the reason is genetic (mutation or chromosome abnormality), but there can be environmental influence (like teratoger

当然，用户也可以进行其它任何形式的语料库检索——单词、词组、子字符串、同义词、搭配词等——然后将检索限制在某一特定的虚拟语料库中。这样的话，虚拟语料库就如同“**语料库中的语料库**”，这对研究某特定话题的人员应该更加有用。English-Corpora.org 与其它语料库网站不同的是，用户仅需轻点几次鼠标再等上几秒钟就能创建虚拟语料库了。

为语言学习者和教师准备的工具（[返回开头](#)）

前文展示的许多检索结果都能为英语学习者和教师提供有用的信息。简洁的**频数表**有时能帮助学生“校正”自己写作某一体裁时的用语。例如，学习者光靠直觉可能不知道 *a lot of NOUN* 这个词组听起来非常不正式，在学术写作中很不常用，而 *several NOUN* 用在正式写作中就恰当得多：

a lot of NOUN								several NOUN							
BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
31758	22679	31043	82391	11017	19551	27189	3537	21535	24730	4023	15315	17372	26952	29037	30919
128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
246.93	182.52	242.38	653.19	93.11	155.05	223.33	29.53	167.44	199.03	31.41	121.42	146.82	213.75	238.51	258.11

正如前面所说，在特定语境下从若干“**相互竞争**”的单词中找出最常用的那个是很有用的，比如在 *argument* 前的与 *powerful* 同义的搭配词。这类知识可以通过超长时间接触第二语言学到，或者，在语料

库中进行几秒检索也能获得，而这样的数据对将英语当作第二语言进行写作的人（包括各个学术领域的研究人员）来说十分珍贵。

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/> STRONG ARGUMENT	331	83	57	3	54	8	38	25	63
2	<input type="checkbox"/> CONVINCING ARGUMENT	218	55	44	12	16	13	21	23	34
3	<input type="checkbox"/> POWERFUL ARGUMENT	148	19	20	2	28	4	17	17	41
4	<input type="checkbox"/> PERSUASIVE ARGUMENT	137	21	23	12	16	5	15	14	31
5	<input type="checkbox"/> EFFECTIVE ARGUMENT	39	6	7	2	12		5	2	5
6	<input type="checkbox"/> POTENT ARGUMENT	12	1	4		2		2	2	1
7	<input type="checkbox"/> FORCEFUL ARGUMENT	13	3	4		1		1	1	3
8	<input type="checkbox"/> VIGOROUS ARGUMENT	10		2	1			1		6
9	<input type="checkbox"/> INFLUENTIAL ARGUMENT	7		1				1		5
	TOTAL	915	188	162	32	129	30	101	84	189

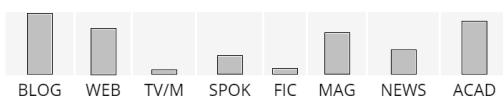
除了前面显示的多种搜索方式之外，本语料库还专为语言学习者设计了一些特色检索功能，在其它任何大型语料库中绝对找不到。例如，在 COCA 和 iWeb 中，用户可以分别浏览这两个语料库中排名前 60000 的单词列表（它们是仅有的最大的、并且经过仔细校正的英语频数列表）。下面词表片段截图展示了三个不同频数段的样本：分别排位在 5000（即该语料库中最常见单词的第 5000 位），25000 和 45000 左右的单词。每个单词都有一个链接，可访问它的“主页”，后面几个图标分别对应发音音频、视频、图像以及翻译（下图）。

2	5197	11377	blogger	NOUN				
3	5198	11374	utterly	ADV				
4	5199	11372	trouble	VERB				
5	5200	11368	texture	NOUN				
6	5201	11365	head	ADJ				
9	25203	576	ergonomic	ADJ				
10	25204	576	tailgate	VERB				
11	25205	576	gasket	NOUN				
12	25206	576	reopening	NOUN				
13	25207	576	impolite	ADJ				
12	45213	113	monotonically	ADV				
13	45214	113	arithmetical	ADJ				
14	45215	113	apolipoprotein	NOUN				
15	45216	113	muddied	ADJ				
16	45217	113	benchmark	VERB				

语料库中排名前 60000 的每个单词（词元）都有各自的“主页”，有这个词（词元）非常丰富的信息，包括：频数、单词排位（如第 1 位—第 60000 位）、按体裁检索的频数、定义，以及多个链接，通过点击相应图标可以分别定位到其它网站的定义、在线词源信息、图片、视频、翻译（超过 100 种语言）、相关话题、搭配词、同义词、词簇（两词、三词或四词组成）、使用该词最多的文本和索引行样本。

(NOUN)

#1487



DMOCGE

 [PlayPhrase](#) [YouGlish](#) [Yarn](#)

IA: Google WordRef Reverso Linguee

SYNONYMS (more)

atmosphere atmosphere, climate, environment, feeling, mood,
sense, situation, surroundings weather environment,
microclimate, temperature, weather

[CLUSTERS](#) (more)

climate •	climate change • climate science • climate scientists • climate in • climate models • climate system • climate change • climate changes
• climate	on climate • to climate • global climate • political climate • about climate • in climate • for climate • with climate
climate ••	climate change in • climate change on • climate change will • climate change has • climate change impacts • climate change to • climate change may • climate change as
•• climate	panel on climate • in the climate • effects of climate • impacts of climate • to the climate • on the climate • in a climate • in this climate
climate •••	climate change is real • climate change is not • climate change and energy • climate change is n't • climate change and global • climate change is already • climate change is happening • climate change and other

[TEXTS / VIRTUAL CORPORA \(more\)](#)

[BLOG:wattsupwiththat.com](#) • [BLOG:judithcurry.com](#) • [BLOG:wattsupwiththat.com](#) • [WEB:...mateshiftproject.org](#) • [BLOG:judithcurry.com](#) • [ACAD:jamba: J Disaster Risk St.](#) • [BLOG:wattsupwiththat.com](#) • [ACAD:EnvirAffairs](#) • [WEB:...terealityproject.org](#) • [BLOG:wattsupwiththat.com](#) • [WEB:uncsd2012.org](#) • [ACAD:Environment](#) • [WEB:aip.org](#) • [WEB:dailytech.com](#) • [ACAD:EnvironmentalHealth](#) • [BLOG:wattsupwiththat.com](#) • [ACAD:The Fletcher Forum of World Affairs](#) • [BLOG:wattsupwiththat.com](#) • [WEB:...ientificamerican.com](#) • [ACAD:Environment](#) • [BLOG:dailykos.com](#) • [BLOG:skepticblog.org](#) • [WEB:wattsupwiththat.com](#) • [BLOG:wattsupwiththat.com](#) • [BLOG:wattsupwiththat.com](#) •

CONCORDANCE LINES (more)

40	MAG: 2009: MotherJones	last year, more than any other group devoted solely to climate	change	But	there are now also 138 lobbyists representing
41	WEB: 2012: counterpunch.org	. # Number two is demanding action to combat worsening climate	change	The	public is ready for this . Hurricane Sandy (
42	NEWS: 2017: USA TODAY	10' glass aquarium and viscerally connects everyday actions to climate	change	(Photo	: Robert Deutsch , USA TODAY) #
43	BLOG: 2012: usnews.nbcnews.com	risk things , esp since NY never had this stuff before climate	changes	have forever	changed NY and NJ (our gov announced that
44	NEWS: 2019: Minneapolis Star Tri...	years , and that trend is projected to continue as the climate	changes	The	pattern's frequency and duration have in fact
45	SPOK: 2002: NPR Science	we come back , can we head off global warming 's climate	changes	We	'll talk about that with someone who thinks we
46	ACAD: 2010: ForeignAffairs	's global population is 6.83 billion .) Barring a cataclysmic climate	crisis	or	a complete failure to recover from the current
47	BLOG: 2012: cameronneylon.net	I should have realised that this would most likely be around climate	data	#	Today the Times reports on its front page that
48	BLOG: 2012: theoil drum.com	over a million worldwide , and contributes to the potential climate	disaster	we face	. The fee for Price Anderson is independent of
49	NEWS: 2019: Minneapolis Star Tri...	. " # Nearing a Tipping Point ? UK Declares " Climate	Emergency	Quartz	has details : " Following the days-long
50	NEWS: 2011: Denver	# It also is important to encourage and cultivate a business climate	for	Colorado companies	, large and small , to purchase products
51	ACAD: 2012: AmJPubHealth	research literature to indicate the importance of work safety climate	for	occupational safety	in agriculture , particularly as

单词“主页”上所有版块的内容都只是概览，用户可以点击几乎任一版块获取**更多信息**。例如，点进动词 *break* 的“词典”页面（该词在 COCA 或 iWeb 中一共有七个页面），就会看到同义词、不同词形的频数、相关词汇、以及语义更具体或更概括的单词。用户可以点击网页上的任一单词，从而转跳到该单词的“主页”。换句话说，所有的单词都是链接在一起的，这样用户就能通过关联词汇追踪单词的“语义路径”。

<

最后，COCA 中的“分析文本”功能为语言学习者和教师提供了诸多非常有用的配置。用户可以输入整篇文本（可以是自己写的，或是来自在线报纸或杂志的文章）。然后，语料库会对文本中的低频词（这些单词往往也是学习者不认识的单词）进行**高亮标识**，并且显示文本中的单词在该语料库中不同频数段的百分比。同时，会呈现不同频数段中所包含的具体单词，并按频数降序排列，这对提取**文本中主题词**很有用。例如，下面**来自 CNN 的文章**（关于“识别 COVID-19 的载体”），文章中的关键主题词分别是 *infected*, *infection*, *antigen*, *symptoms* 和 *virus*。

EDIT TEXT	SAVE TEXT	WORD	PHRASE	(CLICK ANY WORD FOR FULL WORD SKETCH)		
FREQ RANGE	1-500	501-3000	> 3000	LOW FREQ	MID FREQ	HIGH FREQ
1651 WORDS	59 %	11 %	15 %	10: infected 7: infection 6: antigen 5: symptoms, testing-only, virus 4: coronavirus, false, mask 3: baeten, masks, staffers 2: antibody, asymptomatic, fundraiser, infectious, quarantine, rapid, rarely, reagents, sensitive 1: accurate, adviser, asymptomatics, attendees, auditorium, authorization, balcony, bother, cannot, ceremony, cheaper, circulation, comparatively, confirmed, confused, consumables, converting, criticized, crux, czar, dean,	9: strategy 8: failure 7: positive 5: staff 4: negative 3: especially, fail, measures, quickly, wearing 2: ahead, alone, couple, data, distancing, doctors, event, events, everybody, gold, lady, negatives, perfect, personal, professor, recent, safe, seven, standard, true 1: according, active, administration, admitted, advantages, agreed, anybody, anyway, apparently, associates, attended, available, basic, caught, chain, chemicals, circle, complete, completely, contacts, contained, containing, debate, developed,	84: the 40: of 39: and 38: a 31: in 30: to 24: it 21: said, you 20: not 19: they 18: for, testing 15: have, test 14: he, is 13: at, n't, people, tests, that 12: with 11: be, can 10: day, was, which 9: but, every, tested 8: are, as, his, other 7: do, just, time 6: before, did, from, when
CLICK ON ANY WORD BELOW FOR A FULL WORD SKETCH Until President Trump's coronavirus infection , the White House strategy for keeping him and others in the administration safe was one of testing only . The President was rarely seen engaging in two of the most effective and widely promoted public health measures , social distancing and wearing a mask , and many of those who surround him followed his lead . For example, during the recent presidential debate in Cleveland, Trump not only mocked his Democratic rival Joe Biden for wearing a mask , his wife and grown children removed their masks after they were seated in the auditorium , in violation of the events rules . No masks and no back up measures : How the White House became ripe for an outbreak Testing, however, was apparently a strategy Trump could get behind, and so he and his staff were tested often -- the President was said to be tested as often as once a day, possibly more , according to initial reports . But Trump himself admitted earlier this summer he wasn't tested every day. And the White House has not said publicly when the last time the President tested negative before he developed symptoms and tested positive Thursday night . Testing-only strategy a 'complete failure' Unlike mask-wearing , testing would not " send the wrong message " as Trump has said in						

用户可以点击文本中的任一单词，或文本频数列表中的任一单词，就可以查看该词的完整条目，如前所述。这是 COCA 语料库独有的特色，能够让用户轻松**浏览不熟悉单词**，并查看其详细信息。

最后，用户可以点击文本中的任何单词生成短语，然后就能在 **COCA** 中**快捷地找到相关短语**。例如，这篇文章中的短语 *infectious diseases*，用户可点击这两个词（左下图），然后点击“POS”（词性），来检索除 *infectious* 之外的任何形容词，然后点击“FORMS”（词形）来查找 *diseases* 的任何形式（右下图）。

FIND MATCHING PHRASES IN COCA (HELP)

infectious

diseases

+ ADD

- DELETE

*

FORMS

SYNONYMS

POS: NOUN VERB ADJ ADV PREP Other:

SUBMIT

RESET

HELP

FIND MATCHING PHRASES IN COCA (HELP)

ADJ

DISEASE

+ ADD

- DELETE

*

FORMS

SYNONYMS

POS: NOUN VERB ADJ ADV PREP Other:

SUBMIT

RESET

HELP

点击“SUBMIT”（提交）之后，可以在 COCA 中找到相匹配的短语，并按不同体裁中的频数排序。

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	20
1	<input type="checkbox"/> INFECTIOUS DISEASES	1971	121	208	41	192	23	432	212	742	139	220	230	266	
2	<input type="checkbox"/> CARDIOVASCULAR DISEASE	1771	119	251	9	87	3	573	98	631	88	168	223	180	
3	<input type="checkbox"/> INFECTIOUS DISEASE	1300	103	120	63	140	29	240	160	445	92	152	173	184	
4	<input type="checkbox"/> OTHER DISEASES	965	95	145	20	153	17	253	137	145	135	120	168	112	
5	<input type="checkbox"/> CHRONIC DISEASE	930	87	125	10	53	9	174	59	413	39	57	97	101	
6	<input type="checkbox"/> CHRONIC DISEASES	837	79	153	8	53	2	221	55	266	36	53	98	129	
7	<input type="checkbox"/> TRANSMITTED DISEASES	771	45	78	37	133	10	160	119	189	161	121	131	100	
8	<input type="checkbox"/> PULMONARY DISEASE	392	11	52	4	7	1	41	27	249	16	28	28	44	
9	<input type="checkbox"/> CELIAC DISEASE	540	208	153	5	16		109	30	19	1	2	21	43	
10	<input type="checkbox"/> AUTOIMMUNE DISEASE	408	75	64	22	34	3	126	31	53	21	24	51	62	
11	<input type="checkbox"/> AUTOIMMUNE DISEASES	410	67	87	4	20	2	150	26	54	27	26	65	29	
12	<input type="checkbox"/> TRANSMITTED DISEASE	338	31	38	30	77	7	57	40	58	56	56	51	38	
13	<input type="checkbox"/> RESPIRATORY DISEASE	288	15	37	5	10	4	44	42	131	30	22	26	27	
14	<input type="checkbox"/> DEADLY DISEASE	280	19	38	13	66	8	73	39	24	35	26	61	40	

对于教授写作或是用英语写作的非母语研究人员来说，可以“点击即查看”相关短语，特别方便。例如，他们可以点击其文章中的任何短语，查看该短语在不同体裁中的频数（比如判断它是否正式），并且便捷地找到更好的短语（比如如前所述，找到与 [powerful argument](#) 相关的短语）。

其它工具及功能（[返回开头](#)）

如前所述，用户可以进行广泛检索，但刚开始，可能会有些不知所措。幸好每个页面都有一系列“上下文相关”的帮助文档，可以引导用户进行选项设置（例如下图的“搭配词”）。大部分这些帮助文档都提供了示例搜索，通过点击，用户可以先了解熟悉语料库的使用。

List Chart **Collocates** Compare KWIC

Word/phrase [POS] ?

Collocates [POS]

+ 4 3 2 1 0

0 1 2 3 4 +

Find collocates

Reset

☐ Sections Texts/Virtual Sort/Limit Options

(HIDE HELP) **LOGGED IN**

COLLOCATES display: direction/distance

+ 4 3 2 1 0

0 1 2 3 4 +

Select the "span" (number of words to the left and the right) for the collocates. Use + to search more than four words to the left or right, and 0 to exclude the words to the left or right. If you don't select a span, it will default to 4 words left and 4 words right.

The direction of the collocates and the length of the "span" between the "node word" and the collocates is quite important. For example, consider the collocates of *gap* (n): 4 words to the left, 2 words to the right, 4 words to the right, and 4 words left / 4 words right.

此外，每个“结果”页面都有一个[HELP]链接，可以帮助用户理解数据的具体含义：

ON CLICK: [CONTEXT] [ALL] [BLC]

HELP

CONTEXT

1

2

3

4

SOFT TISSUE

SOFT DRINK

1120

1120

62

62

Note: these are the partial results for *soft* + NOUN in COHA. Another search (in another corpus) will of course yield different results, but the general concepts remain the same.

WORD	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	SUB	TOT
1 SOFT DRINKS	5	20	31	39	60	42	38	27	32	65	294
2 SOFT MONEY		4						45	30	45	79
3 SOFT DRINK	6	8	10	16	20	16	23	10	5	33	114
4 SOFT VOICE	7	5	12	7	6	2	14	4	6	18	63

- The rank-ordered list of words or phrases in the results set. Click on the word or phrase to see the "Keyword in Context" display, with all entries for this word or phrase in all decades.
- These columns show the frequency of the word or phrase in each decade from the 1920s-2000s. If you have selected a particular century or register in Section 1 of the search interface, the selected columns will be highlighted in the results set.
- If you have selected a particular decade (or set of decades) in Section 1 of the search interface, then this column will show the total number of hits for each word or phrase.

用户可以查看他们的检索“历史”，甚至可以找到以往检索过的特定单词或短语，然后复制检索的链接，将其嵌入到研究论文或是网页中，这样其它人就能看到他们最初的检索内容（从而使语料库的检索结果“可复制”）。

<div> <div></div> <div>SUBMIT</div> <div>HELP</div> </div> <div> <input checked="" type="radio"/> CORPUS SEARCH <input type="radio"/> MY NOTES </div>	<div>CORPORA USED (LAST 6 MONTHS)</div> <div> <div>COCA</div> <div>1999, CORONA 591, NOW 403, IWEB 176, GLOWBE 101, COHA 97, TIME 67, CAN 48, BNC 30, GC 29, TV 18, WIKI 13, EEBO 9, GOOGLE-SP 4, HANS 3, CORE 2, SOAP 1, MOVIES 1</div> </div>
---	---

HIDE	<p>Copy the following web address into a web page, email, or other document, to see the same results from the corpus as when you did the search yourself.</p> <p>https://www.english-corpora.org/glowbe/?c=glowbe&q=92716465</p>								
HELP	ADD NOTE	HIDE	RE-DO	SHARE LINK	CORPUS	WORD(S)	SECTIONS	TYPE	WHEN
1					GLOWBE	CONJ PRON BE like ,		CHART	10/29/2020
2					COCA	CONJ PRON BE like ,		CHART	10/28/2020
3					COCA	VERB likely VERB		CHART	10/28/2020
4					COCA	BE likely the		CHART	10/28/2020
5					COCA	CONJ PRON BE like ,		CHART	10/27/2020
6					COCA	CONJ PRON BE like ,		TABLE	10/27/2020

通过添加注释或评论的方式，用户可以“标注”其检索，然后在这些标注中检索所有匹配项（例如，检索某一课程讲座或是其正在撰写的论文）。

HELP	EDIT NOTE	HIDE	RE-DO	SHARE LINK	CORPUS	WORD(S)	SECTIONS	TYPE	WHEN
1					COCA	CONJ PRON BE like ,		CHART	10/27/2020
	<div> <div>Presentation on syntactic change in <u>ELang</u> 325</div> <div>ADD NOTE FOR QUERY (Note: remove note above to delete it)</div> </div>								

用户也可以保存索引行，并划分成不同类别（注意下图中的三种不同颜色）：

CLICK FOR MORE CONTEXT					<input type="checkbox"/>	<input checked="" type="checkbox"/> [?] <div>SAVE LIST</div> CHOOSE LIST <div>.....</div> CREATE NEW LIST <div>soft_voice</div> [?]	SHOW DUPLICATES
1	2012	WEB	rhrealitycheck.org	A	B	C	. He was perfectly bald, with thick glasses, and wooden clogs, a soft voice . # A squirt of blue gel on my belly for the fetal monitor
2	2016	FIC	Analog	A	B	C	An expected response. " Still looking at her calmly, the man raised his soft voice : " Captain Pinkerton, if you please. " # She turned,
3	1994	SPOK	ABC_Nightline	A	B	C	Greenwood City Council: Bob is one of those persons or individuals with a very soft voice , very intelligent, and very easy going. And he won a lot
4	2012	WEB	academyofbards.org	A	B	C	'm leaving the agenda for Monday's meeting just went to Development, " a soft voice behind her announced, mercifully interrupting her introspection
5	2008	FIC	Triquarterly	A	B	C	So then I'm walking out the room and I hear Trudy saying in this soft voice , " Dave's kind of tired. Long flight. " I get
6	2012	BLOG	...ppinbob.blogspot.com	A	B	C	was inspired after the couple was having difficulty communicating by telephone. Audrey had a soft voice and was unable to speak up so her husband
7	2005	FIC	NewYorker	A	B	C	the picture under the naked bulb of his room, he said, in a soft voice , " I took him here to Xian for his graduation. To sightsee
8	2002	FIC	VirginiaQRev	A	B	C	body lay; she hadn't seen him enter the room. He heard a soft voice say, " There's Brian, " and then another one, not

之后，还可以对这些索引行进行扩充、删除和移动：

<input type="checkbox"/> # <input type="text"/> <input type="button" value="DELETE ENTRIES"/> <input type="button" value="MOVE ENTRIES"/> <input type="button" value="EXPAND ENTRIES"/> <input type="button" value="F7"/>		
1	<input type="checkbox"/> COCA:1994:SPOK ABC_Nightline	Bob Moses's work to empower blacks was known across Mississippi. He was jailed, he was beaten, but his quiet manner touched many and turned them up the road to freedom. DAVID JORDAN, President, Greenwood City Council: Bob is one of those persons or individuals with a very soft voice , very intelligent, and very easy going. And he won a lot of people over, and people attempted to register to vote, and when they attempted to register to vote in this town, dogs were put on them. And the few that registered to vote, their names
2	<input checked="" type="checkbox"/> COCA:2012:WEB academyofbards.org	, certain, a hint of aggression she inspired confidence. Michael sighed, and hoped her assessment of the woman she had hired the day before was correct. She was going to need help. # "I'm leaving the agenda for Monday's meeting just went to Development," a soft voice behind her announced, mercifully interrupting her introspections. # Michael swiveled away from the window to face the door. She smiled tiredly at the brunette in the doorway. "Yes, fine. Thank you." # "I will, soon," Michael lied, appreciating the concern
3	<input type="checkbox"/> COCA:2002:FIC VirginiaQRev	into the room and stood by the door for a moment. There were several of his aunts and uncles and cousins gathered around his grandmother, who was sitting in a chair by the bed where his grandfather's body lay; she hadn't seen him enter the room. He heard a soft voice say, "There's Brian," and then another one, not so soft, said, "Late as usual." # A nurse entered the room a few seconds after Brian and his mother, before the door had completely closed. The nurse stood against the wall by
4	<input type="checkbox"/> COCA:2012:WEB rrealitycheck.org	; they moved so swiftly, with such seriousness. # I had a new doctor now. Lisa, obstetrician of the normal, was instantly off my case, and I was assigned a special neonatologist named Weiss. He was perfectly bald, with thick glasses, and wooden clogs, a soft voice . # A squirt of blue gel on my belly for the fetal monitor, the galloping sound of hoof beats, the baby riding a wild pony inside me. What a relief to hear that sound, although I did n't need the monitor; I could feel the baby punching at

用户如果想检索一系列单词，例如与身体、情绪或是某类动词相关的一组单词，还可以创建“定制词表”。

MARK_DAVIES@BYU.EDU

To modify a list:
Add or delete words from your list and click on [Submit]
[Return to main menu](#)
[Help](#)

1-WORD	<input type="button" value="M"/>	<input type="button" value="D"/>	ING1	<input type="button" value="M"/>	<input type="button" value="D"/>
10-WORDS	<input type="button" value="M"/>	<input type="button" value="D"/>	INTO-SEE	<input type="button" value="M"/>	<input type="button" value="D"/>
200-WORDS	<input type="button" value="M"/>	<input type="button" value="D"/>	INTO2	<input type="button" value="M"/>	<input type="button" value="D"/>
99-WORDS	<input type="button" value="M"/>	<input type="button" value="D"/>	JUSTICIA	<input type="button" value="M"/>	<input type="button" value="D"/>
BEAUTIFUL-SYN	<input type="button" value="M"/>	<input type="button" value="D"/>	MILITARY	<input type="button" value="M"/>	<input type="button" value="D"/>
BODY	<input type="button" value="M"/>	<input type="button" value="D"/>	MIRAR	<input type="button" value="M"/>	<input type="button" value="D"/>
BONITO-PORT	<input type="button" value="M"/>	<input type="button" value="D"/>	MUJER	<input type="button" value="M"/>	<input type="button" value="D"/>

MODIFY LIST

WORDLIST NAME

body

LIST OF WORDS

head
face
back
arm
hair
leg

SUBMIT RESET

然后将这些词表打包使用（如下图的“@body”，指定定制的与身体有关的词表），从而“在语义层面”对语料库进行检索：

List Chart Word Browse +

sore @body

[POS]?

Find matching strings

Reset

☒ Sections
 ☐ Texts/Virtual
 ☐ Sort/Limit
 ☐ Options

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/> SORE THUMB	236	66	30	27	21	36	28	22	6
2	<input type="checkbox"/> SORE BACK	140	16	13	9		17	31	53	1
3	<input type="checkbox"/> SORE SHOULDER	83	2	8	3	2	8	12	47	1
4	<input type="checkbox"/> SORE KNEE	57	8	4	2	2	9	7	25	
5	<input type="checkbox"/> SORE ARM	45	5	4	1	4	11	3	16	1
6	<input type="checkbox"/> SORE NECK	44	9	4	2	3	7	7	12	
7	<input type="checkbox"/> SORE ANKLE	34	3	2	1	1	2	2	23	
8	<input type="checkbox"/> SORE ELBOW	27	1	1		1	1	1	21	1
9	<input type="checkbox"/> SORE FOOT	27	1	6	1	3	3	6	7	
10	<input type="checkbox"/> SORE HEAD	23	2		7	1	12	1		

在任何检索“结果”页面，都有着丰富的外部资源链接，例如翻译（超过 100 多种语言），谷歌搜索的网页、图像和书籍，以及单词的发音和视频。

ON CLICK:

HELP	CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	<input type="checkbox"/> SOFT TISSUE	1120	62	67	74	39	36	100	35	707
2	<input type="checkbox"/> SOFT DRINKS	1109	123	123	42	90	83	304	296	48
3	<input type="checkbox"/> SOFT MONEY	790	21	34	12	446	8	78	153	38
4	<input type="checkbox"/> SOFT SPOT	867	133	110	166	63	159	135	86	15
5	<input type="checkbox"/> SOFT DRINK	721	43	60	48	77	68	191	199	35
6	<input type="checkbox"/> SOFT VOICE	546	12	39	10	11	351	53	52	18
7	<input type="checkbox"/> SOFT POWER	421	49	53	1	64	1	51	32	170

GERMAN ENGLISH SPANISH

alkoholfreie Getränke

最后，研究人员还可以下载大量基于在线语料库的数据，方便离线时使用。例如，下载全文数据 (www.corpusdata.org)，频数数据 (www.wordfrequency.info)，搭配词 (www.collocates.info) 以及 N 元组 (www.ngrams.info)。

小结

English-Corpora.org 的系列语料库集是**世界上使用最广泛的语料库**，每月用户超过 13 万人次，包括研究人员、教师和语言学习者。这些语料库已成为每年数以千计的研究**论文**的写作基石，也是世界各地**课堂**教学中不可或缺的一部分。

在 English-Corpora.org，研究人员可以观察到**英语的变体**（如基于体裁的、历时的以及方言的），而在其它语料库几乎无法实现。历史、文化以及法律等特定领域的研究人员还可以利用语料库中的大量文本来研究**社会文化问题**。这些语料库为（**广泛学术领域**）的**非母语研究人员**的英语写作提供分析工具，是标准词典和同义词词典都无法匹及的。同时，English-Corpora.org 还为**语言学习和教学**提供了丰富的可能，让它独一无二。

我们感谢以下人员，他们帮助将本文档翻译成普通话。

[愿风裁尘](#) (Derek Lee): 李志豪 黑龙江大学

[黄锦鸿](#) (Jinhong Huang): 复旦大学外文学院在读博士，研究方向：词典学、词汇学、术语学

[谭宇琪](#) (Yuqi Tan): 谭宇琪，先后毕业于北京师范大学和香港中文大学（深圳），现为全职翻译

[三合泥](#) (Ting LUO): 罗婷，英语教师，现任职于中国民用航空飞行学院

[郭顺](#) (Guo Shun): 浙江科技学院

[归雁阔客](#) (Zhilong Xi): 席志龙，电子科技大学硕士研究生，主要研究方向为强化学习和无人机控制，对语言学也十分感兴趣

[Qian Shen](#)