# INGLÊS

English around the world: o uso da língua em interações e comunicações interculturais

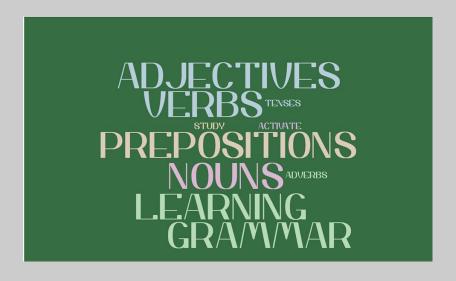
## Língua em uso: padrões

### **OBJECTIVE:**

### **Explore language in use:**

- Lexicogrammar
- Patterns
- Collocations

**Grammar** 



### **Lists of vocabulary**



How do we combine them?
When / Where / How do we use them?

- 1) pay...
- 2) be...
- 3) speak...
- 4) do
- 5) avoid

- ( ) using slang
- ( ) careful
- ( ) slowly
- ( ) research
- ( ) attention

- 1) Pay attention...
- 2) Be careful...
- 3) Speak slowly
- 4) Do research...
- 5) Watch the humour
- 6) Avoid using slang

- 1) Pay attention...
- 2) Be careful...
- 3) Speak slowly
- 4) Do research...
- 5) Watch the humour
- 6) Avoid using slang

Pay...

What do you pay?

### Corpus of Contemporary American English 🕞 🚯 🛈 🖹 🕜 😲 🔱

**SEARCH** 









CONTEXT







**TEXTS** 

#### **FREQUENCY**

**DOWNLOAD LIST OF ALL 485,179 TEXTS AND SUMMARY BY YEAR, GENRE, AND SUB-GENRE** 

The corpus is composed of more than one billion words in 485,202 texts, including 20 million words each year from 1990-2019. For each year (and therefore overall, as well), the corpus is evenly divided between the genres of TV and Movies subtitles, spoken, fiction, popular magazines, newspapers, and academic journals.

YEAR	BLOG	WEB	TV / MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	TOTAL
	125,496,215	129,899,426	128,013,334	127,396,916	119,505,292	127,352,014	122,959,393	120,988,348	1,001,610,938
1990			3,207,900	4,374,469	4,162,242	4,101,447	4,082,931	3,983,143	23,914,122
1991			3,379,151	4,316,898	4,192,646	4,209,838	4,104,806	4,051,046	24,256,376
1992			3,183,858	4,523,054	3,893,956	4,288,694	4,092,031	4,028,147	24,011,732
1993			3,785,924	4,487,978	3,921,244	4,254,351	4,153,070	4,150,671	24,755,231
1994			4,375,338	4,457,726	3,870,757	4,310,375	4,147,947	4,047,115	25,211,252
1995			5,006,966	4,548,602	3,846,412	4,314,737	4,122,703	4,016,371	25,857,786
1996			4,384,976	4,095,266	3,758,787	4,338,766	4,099,305	4,110,209	24,789,305
1997			4,380,670	3,904,996	3,617,741	4,368,917	4,153,906	4,420,786	24,849,013
1998			4,390,197	4,446,217	3,779,801	4,393,835	4,122,295	4,111,453	25,245,796
1999			4,381,144	4,445,564	4,154,537	4,391,146	4,107,423	4,023,282	25,505,095
2000			4,385,593	4,455,815	3,942,474	4,387,935	4,037,086	4,093,991	25,304,894
2001			4,389,164	4,026,240	3,894,789	4,298,636	4,072,447	3,965,654	24,648,931
2002			4.384.475	4.372.290	3.766.673	4.310.634	4.114.280	4.054.359	25.004.713



#### **Corpus of Contemporary Ameri**















**FREQUENCY** SEARCH CONTEXT **TEXTS** 

- TV/Movies subtitles: (128 million words [128,013,334]). These come from the American part of the TV and Movies corpora. These subtitles are as informal (or more informal) than actual spoken data. The texts were taken from the OpenSubtitles collection. In cases where there were multiple subtitles files for a given TV episode (which was the norm), we used the "highest ranked" file, in terms of accuracy (from the ratings at OpenSubtitles). We then matched up each episode (and series) with the corresponding page from IMDB, which provides rich metadata for these TV shows (and which can be used to create your own Virtual Corpus).
- Spoken: (127 million words [127,396,916]) Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Today Show (NBC), 60 Minutes (CBS), Hannity and Colmes (Fox), Jerry Springer, etc). [See notes on the naturalness and authenticity of the language from these transcripts).
- Fiction: (120 million words [119,505,292]) Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books 1990-present, and movie scripts.
- Popular Magazines: (127 million words [127,352,014]) Nearly 100 different magazines, with a good mix (overall, and by year) between specific domains (news, health, home and gardening, women, financial, religion, sports, etc). A few examples are Time, Men's Health, Good Housekeeping, Cosmopolitan, Fortune, Christian Century, Sports Illustrated, etc.
- Newspapers: (123 million words [122,959,393]) Ten newspapers from across the US, including: USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. In most cases, there is a good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.
- Academic Journals: (121 million words [120,988,348]) Nearly 100 different peer-reviewed journals. These were selected to cover the entire range of the Library of Congress classification system (e.g. a certain percentage from B (philosophy, psychology, religion), D (world history), K (education), T (technology), etc.), both overall and by number of words per year.
- Blogs: (125 million words [125,496,215]). These texts represent a subset of the texts from the United States in the GloWbE corpus. At that time, Google allowed searches to be restricted to blogs, so nearly all of these texts are actually blogs.
- Web pages: (130 million words [129,899,426]). They represent a subset of the "General" texts from the United States in the GloWbE corpus. Some of these texts are actually blogs (there was no way to search "NOT blogs" in Google at that time).
- -- More information on how these texts were obtained from Google searches, and how the texts were identified as being from the US.
- -- For both blogs and general web pages, these were subsequently categorized by Serge Sharoff, so that in COCA you can limit searches to a particular web genre.
- -- Note that these texts were all collected in Oct 2012, so they are more of a "snapshot" of this genre, rather than year by year (as above). As a result, they are not included in the "historical" data, when you compare the frequency across decades or year. All historical data comes from the other six genres listed above.



### **English-Corpora.org**

home corpora users related resources my account upgrade help

The most widely used online corpora. Overview, search types, looking at variation, corpus-based resources.

The links below are for the online interface. But you can also 🕔 download the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
iWeb: The Intelligent Web-based Corpus	•	14 billion	6 countries	2017	Web
News on the Web (NOW)	•	10. billion+	20 countries	2010- <mark>yesterday</mark>	Web: News
Global Web-Based English (GloWbE)	•	1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	•	1.9 billion	(Various)	2014	Wikipedia
Corpus of Contemporary American English (COCA)	•	1.0 billion	American	1990- <b>2019</b>	Balanced
Corpus of Historical American English (COHA)	•	400 million	American	1810-2009	Balanced
The TV Corpus	•	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	0	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	•	100 million	American	2001-2012	TV shows
Hansard Corpus		1.6 billion	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced
Strathy Corpus (Canada)		50 million	Canadian	1970s-2000s	Balanced
CORE Corpus		50 million	6 countries	2014	Web
From Google Books n-grams (compare)					
American English		155 billion	American	1500s-2000s	(Various)
British English		34 billion	British	1500s-2000	(Various)





### **English-Corpora.org**

home corpora users related resources my account upgrade help

The most widely used online corpora. Overview, search types, looking at variation, corpus-based resources.

The links below are for the online interface. But you can also 🕔 download the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
iWeb: The Intelligent Web-based Corpus	•	14 billion	6 countries	2017	Web
News on the Web (NOW)	0	10. billion+	20 countries	2010- <mark>yesterday</mark>	Web: News
Global Web-Based English (GloWbE)	0	1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	•	1.9 billion	(Various)	2014	Wikipedia
Corpus of Contemporary American English (COCA)	0	1.0 billion	American	1990- <b>2019</b>	Balanced
Corpus of Historical American English (COHA)	•	400 million	American	1810-2009	Balanced
The TV Corpus	0	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	0	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	0	100 million	American	2001-2012	TV shows
Hansard Corpus		1.6 billion	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced
Strathy Corpus (Canada)		50 million	Canadian	1970s-2000s	Balanced
CORE Corpus		50 million	6 countries	2014	Web
From Google Books n-grams (compare)					
American English		155 billion	American	1500s-2000s	(Various)
British English		34 billion	British	1500s-2000	(Various)



#### Corpus of Contemporary American English 🔓 🚯 🛈 🖹 🕜 😲 🕕





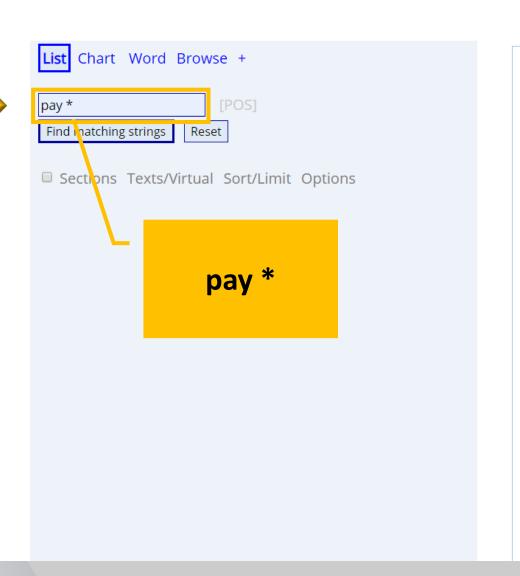








**SEARCH FREQUENCY** CONTEXT **ACCOUNT** 





Take a short tour, including new features in March 2020



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

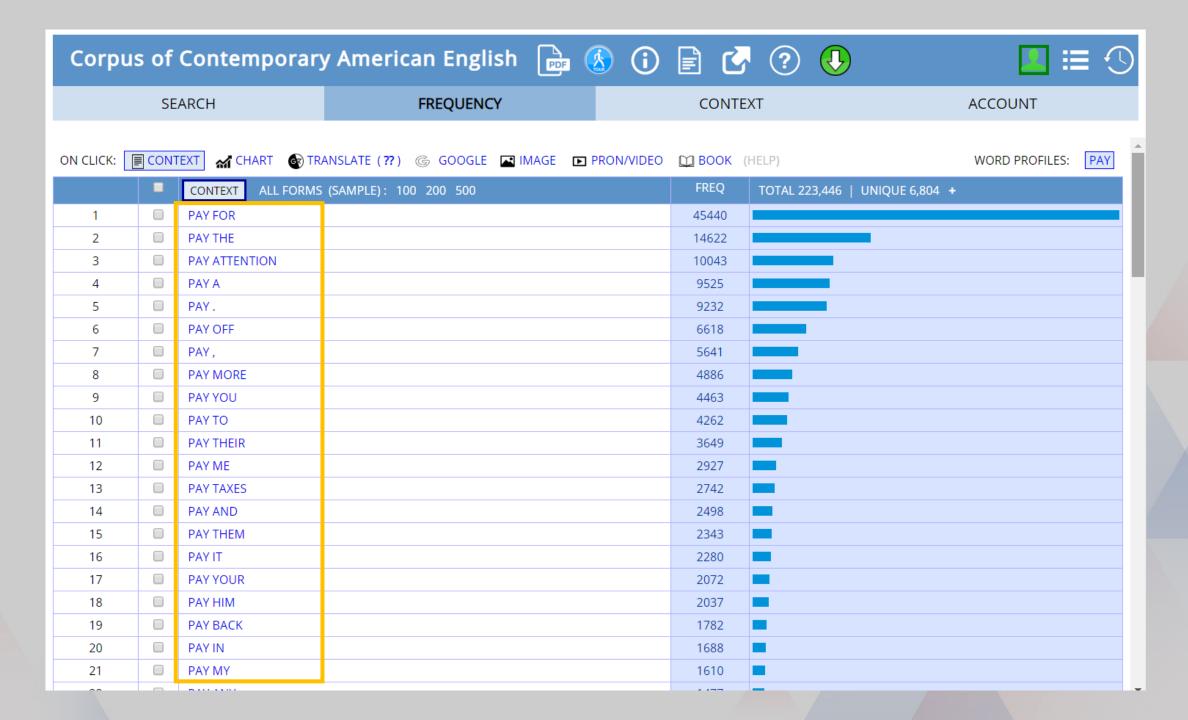
The Corpus of Contemporary American English (COCA) is the only large, genrebalanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English.

The corpus contains more than one billion words of text (20 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.









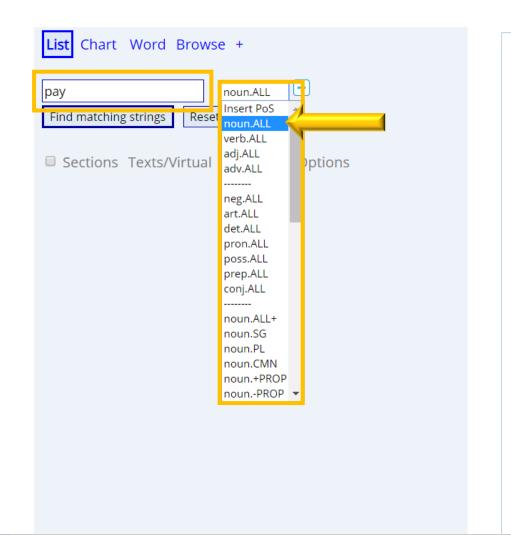








**SEARCH FREQUENCY ACCOUNT** CONTEXT





Take a short tour, including new features in March 2020



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

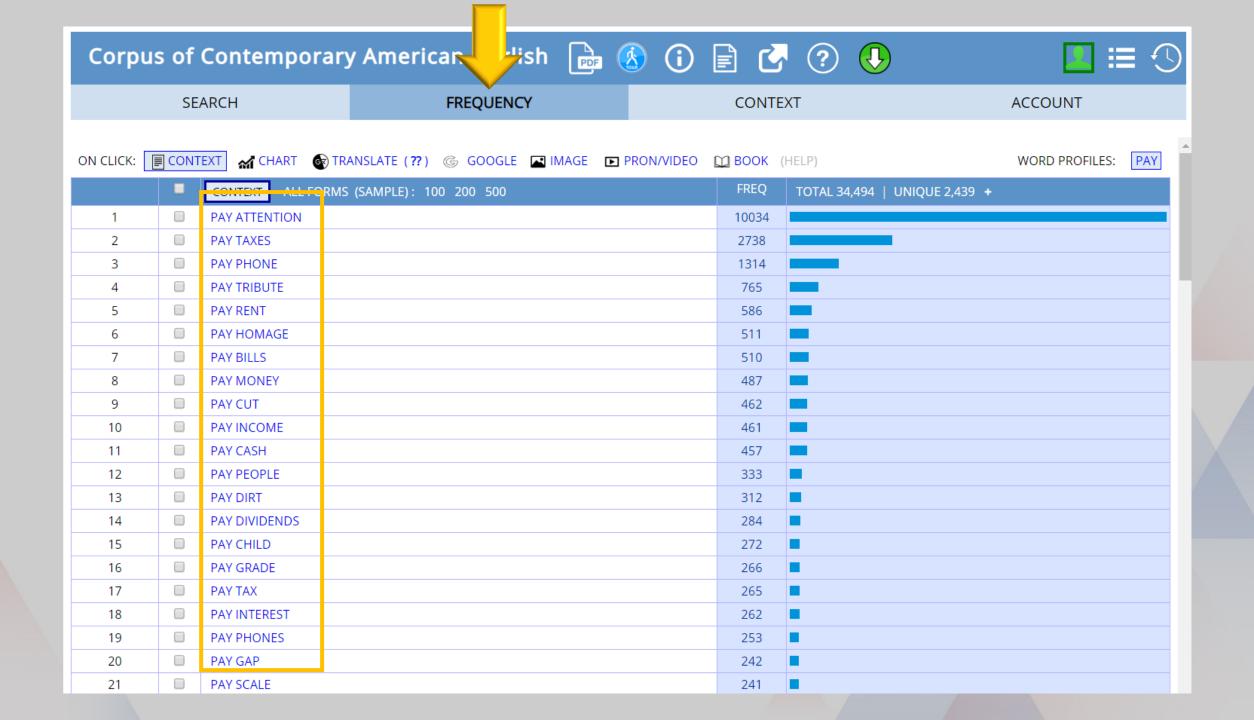
The Corpus of Contemporary American English (COCA) is the only large, genrebalanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English.

The corpus contains more than **one billion** words of text (20 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.



## Corpus of Contemporary American English 🕒 🚯 🛈 🖹











SEARCH FREQUENCY **CONTEXT ACCOUNT** 

FIND SAMPLE: <u>100</u> <u>200</u> <u>500</u> <u>1000</u>

PAGE: << < 1/101 > >>

			CONTEXT		[?]	l	SAVE LIST CHOOSE LIST CREATE NEW LIST [?] SHOW DUPLICATES
1	2012	BLOG	addictinginfo.org		В		assistance, but not the dire number you presented. # You obviously didn't pay attention. That would have gotten a lot worse by now if Obama hadn't
2	2012	BLOG	addictinginfo.org	Α	В	С	) # People who like sports typically really like sports. People who like sports pay attention to which people play for which team. They know the coaches' n
3	2012	BLOG	addictinginfo.org	Α	В	С	the problem: Political issues have a habit of affecting us all, whether we pay attention to them or not. Conversely, whether a particular sports team wins
4	2012	BLOG	dailykos.com	Α	В	С	you someone from Texas should not be president of the United States, please, pay attention. " Molly Ivins # except it would be a frontal assault of
5	2012	BLOG	tionalgeographic.com	Α	В	С	7:29 pm # To the people saying " It's only a movie " people pay attention to the portrayals of animals in movies. It's sad but true the
6	2012	BLOG	coronadocrest.com	Α	В	С	four teenage girls hear about such a thing, it's a little hard to pay attention to anything else. # We chattered about the news, wishing we could
7	2012	BLOG	rlie-bird-parker.com	Α	В	С	hosting provider ads will show up on their website, specifically because they did not pay attention when choosing their package for hosting. In most case
8	2012	BLOG	ssociationmavens.com	Α	В	С	said you need a much stronger commitment to personal development. You need to actually pay attention to letting your employees uncover their own d
9	2012	BLOG	blogs.cgdev.org	Α	В	C	officials (and the professionals who work for them) to see what makes them pay attention to particular think tanks. # At the European Centre for Develop
10	2012	BLOG	obsguy.wordpress.com	Α	В	С	job satisfaction. A job is a terrible thing to waste, choose wisely and pay attention to the fit. You'll get many more years of enjoyment and job
11	2012	BLOG	nstructionlitmag.com	Α	В	С	through the first couple of months of the season while most fans and degenerate gamblers pay attention to the NFL playoff race. This is fine. Between no
12	2012	BLOG	blog.livestrong.org	Α	В	C	I knew also that if I don't speak about cancer no one else could pay attention on cancer. That is why I decided to bring Agir Ensemble becoming UICC
13	2012	BLOG	erblog.thinkport.org	Α	В	C	. Designers often evaluate " risk vs. reward, " whereby the player learns to pay attention to his/her in-game actions to see if it is woth it to do effect
14	2012	BLOG	omicperspectives.org	Α	В	C	with these terms: Kelton teeter-totter (This is one where you do have to pay attention, but it was the breakthrough for me. The big ah-ha.) (
15	2012	BLOG	jihadwatch.org	Α	В	C	the London transit slaughter by Muslims, note how he blames Bush, and also pay attention how he insisted Muslims must Police themselves, yet now Fa
16	2012	BLOG	kateharding.net	Α	В	С	our heads. Usually, people are too busy with their own workouts to even pay attention to how big your ass is or whatever. Once we become comfortable
17	2012	BLOG	antipope.org	Α	В	С	: look for a local writers' workshop. Sit in on a session. Pay attention to the internal social dynamic as much as the fiction under discussion. If
18	2012	BLOG	dailykos.com	Α	В	С	always had current events and civics as well. My parent's always made us pay attention when something big or important was happening in politics, of co
19	2012	BLOG	anantonio.jobing.com	Α	В	C	resume real estate " summation that conveys what you're about and why employers should pay attention. Laura Smith-Proulx, Executive Resume Writer
20	2012	BLOG	themideastupdate.com	Α	В	С	hints and painting a corner of the picture publicly, but one really has to pay attention. The only way to figure out how Israel and the West plan to
					_	_	

## Corpus of Contemporary American English 🔓 🚯 🛈











SEARCH FREQUENCY **CONTEXT ACCOUNT** 

FIND SAMPLE: <u>100</u> <u>200</u> <u>500</u> <u>1000</u>

PAGE: << < 1 / 101 > >>

PAG	PAGE: << 1/101 > >>													
CLIC	K FOR	MORE (	CONTEXT		[?]	SAVE LIS	CHOOSE LIST		CREATE NEW LIST		[?]	SHOW DUPLICATES		
1	2012	BLOG	addictinginfo.org	Α	В	Cassistar	ssistance, but not the dire number you presented. # You obviously didn't pay attention. That would have gotten a lot worse by now if Obama hadn't							
2	2012	BLOG	addictinginfo.org	Α	В	) # Peo	People who like sports typically really like sports. People who like sports pay attention to very hich people play for which team. They know the coaches' n							
3	2012	BLOG	addictinginfo.org	Α	В	the pro	blem: Political issu	es have a h	nabit of affecting us a	ll, whether we	pay attenti	on to hem or not. Conversely, whether a particular sports team wins		
4	2012	BLOG	dailykos.com	Α	В	you sor	neone from Texas	should not	be president of the	United States,	please, <b>pay</b>	attention. " Molly Ivins # except it would be a frontal assault of		
5	2012	BLOG	tionalgeographic.com	Α	В	7:29 pn	n # To the people s	saying " It's	only a movie " peop	e <b>pay attentio</b> r	to the por	trayals of animals in movies. It's sad but true the		
6	2012	BLOG	coronadocrest.com	Α	В	four tee	enage girls hear ab	out such a	thing, it's a little har	d to <b>pay attent</b> i	o i to a nyth	ning else. # We chattered about the news, wishing we could		
7	2012	BLOG	rlie-bird-parker.com	Α	В	C hosting	provider ads will s	show up on	their website, speci	fically because	they did no	ot pay attention when choosing their package for hosting. In most case		
8	2012	BLOG	ssociationmavens.com	Α	В	said yo	u need a much stro	onger comn	mitment to personal	development.	You need t	o actually <b>pay attentio</b> n to etting your employees uncover their own d		
9	2012	BLOG	blogs.cgdev.org	Α	В	C officials	(and the profession	onals who w	work for them) to see	what makes t	hem <b>pay at</b>	tention to particular think tanks. # At the European Centre for Develor		
10	2012	BLOG	obsguy.wordpress.com	Α	В	job sati	sfaction. A job is a	terrible thir	ng to waste, choose	wisely and <b>pay</b>	<b>attention</b> t	o the fit. You'll get many more years of enjoyment and job		
11	2012	BLOG	nstructionlitmag.com	Α	В	C through	the first couple o	f months of	f the season while m	ost fans and d	egenerate (	gamblers <b>pay attention</b> to the NFL playoff race. This is fine. Between no		
12	2012	BLOG	blog.livestrong.org	Α	В	C I knew	also that if I don't s	speak about	t cancer no one else	could <b>pay atte</b>	<b>ntion</b> on ca	ncer. That is why I decided to bring Agir Ensemble becoming UICC		
13	2012	BLOG	erblog.thinkport.org	Α	В	C . Design	ners often evaluate	" risk vs. re	eward, " whereby th	e player learns	to <b>pay atte</b>	ntion to his/her in-game actions to see if it is woth it to do effect		
14	2012	BLOG	omicperspectives.org	Α	В	c with the	ese terms: Kelton t	eeter-totter	r (This is one where	you do have to	pay attent	ion, but it was the breakthrough for me. The big ah-ha.) (		
15	2012	BLOG	jihadwatch.org	Α	В	the Lon	don transit slaugh	ter by Musl	lims, note how he bl	ames Bush, an	d also <b>pay a</b>	attention how he insisted Muslims must Police themselves, yet now Fa		
16	2012	BLOG	kateharding.net	Α	В	our hea	ds. Usually, people	e are too bu	usy with their own w	orkouts to eve	n <b>pay atten</b>	tion to how big your ass is or whatever. Once we become comfortable		
17	2012	BLOG	antipope.org	Α	В	: look fo	or a local writers' w	orkshop. Si	it in on a session. <b>Pa</b>	<b>y attention</b> to t	he internal	social dynamic as much as the fiction under discussion. If		
18	2012	BLOG	dailykos.com	Α	В	always	had current events	s and civics	as well. My parent's	always made u	ıs <b>pay atter</b>	ntion when something big or important was happening in politics, of co		
19	2012	BLOG	anantonio.jobing.com	Α	В	C resume	real estate " sumr	mation that	conveys what you'r	e about and wh	ny employe	rs should <b>pay attention</b> . Laura Smith-Proulx, Executive Resume Writer		
20	2012	BLOG	themideastupdate.com	Α	В	C hints ar	nd painting a corne	er of the pic	ture publicly, but or	e really has to	pay attenti	on. The only way to figure out how Israel and the West plan to		
1				1.1	-									

#### Corpus of Contemporary American English 🕞 🚯 🕡 🖹 🕜 🕐

















CONTEXT + **SEARCH FREQUENCY** CONTEXT

#### Source information:

Source	BLOG http://www.antipope.org/charlie/blog-static/2012/11/an-interim-report-from-the-coa.html
Date	2012
Title	An interim report from the coal face - Charlie's Diary

#### **Expanded context:**

answer it, but my answer would be 12 or more years out of date, and we're in the middle of a gigantic and fast-moving shift in the way the industry works. # One warning: family will either diplomatically ignore your writing or tell you exactly what they think you want to hear. Friends will do likewise unless they're also writers and they're trying to raise their game with an eye to eventual publication and they respect you. My rec would be: look for a local writers' workshop. Sit in on a session. Pay attention to the internal social dynamic as much as the fiction under discussion. If it's 100% mutual ego-stroking, don't go back: it won't help you. If it's a groupthink cult, don't go back, it won't help you. What you're looking for is honest feedback including constructive criticism. # Agents: a good one is worth their weight in, if not gold, then probably silver: they will earn you far more money than they take in their commission

- 1) Pay attention...
- 2) Be careful...
- 3) Speak slowly
- 4) Do research...
- 5) Watch the humour
- 6) Avoid using slang

### Corpus of Contemporary American English 🕞 🚯 😯 🖹 🕜 😲







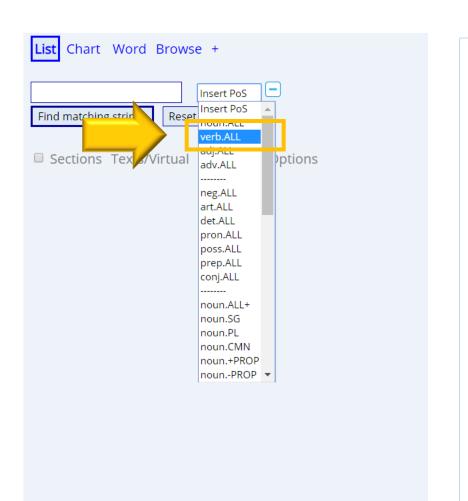








**FREQUENCY SEARCH** CONTEXT CONTEXT +





OGGED IN

Take a short tour, including new features in March 2020



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

The Corpus of Contemporary American English (COCA) is the only large, genrebalanced 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 (20 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.

#### Corpus of Contemporary American English 🔓 🚯 🛈 📄 🕜 😲 🕕

















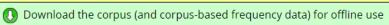
**SEARCH FREQUENCY** CONTEXT CONTEXT +



Sections Texts/Virtual Sort/Limit Options



Take a short tour, including new features in March 2020



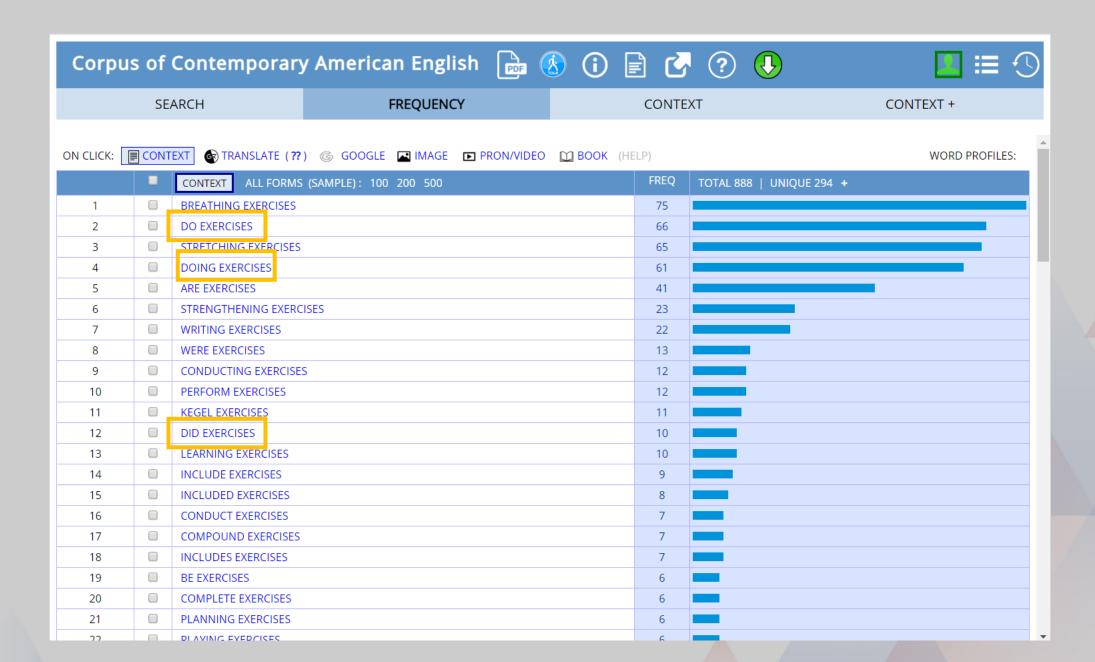
The Corpus of Contemporary American English (COCA) is the only large, genrebalanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English.

The corpus contains more than **one billion** words of text (20 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.



## Língua em uso: padrões

- ✓ Pay attention to patterns.
- ✓ Search for collocations.