

## The NOW Corpus as the most robust monitor corpus of English

The [NOW Corpus](#) from [English-Corpora.org](#) currently has (as of mid-May 2022) about 15.1 billion words of data from 21 English-speaking countries, and it is growing by about 200-220 million words per month (~2.4 billion words per year). **The NOW Corpus is the only resource that allows you to find the frequency of words, phrases, and collocates by year, month, and day – in thousands of newspapers and magazines from throughout the world.**

Tools like [Google Trends](#) allow you to see what people are searching for over time, but not what is actually being used in newspapers and magazines. [LexisNexis](#) has some functionality for showing frequency over time, but it costs thousands of dollars per year. As far as "structured corpora", [MonitorCorpus](#) – a site that used to provide similar functionality – is no longer operational. Oxford University Press apparently has a "monitor corpus" for [internal use](#), but it is not available to the public.

Other than NOW, the only other publicly-available monitor corpus is the "Trends" corpus from [Sketch Engine](#). As of May 2022, it provides about 900 million words of data (mostly from January 2020 to the present). A sample search for *Ukraine* in the Trends corpus is shown below:

English Trends preloaded/rss\_en

English Trends corpus is a monitor corpus consists of news articles gained from RSS links. The co

**GENERAL INFO**

Language: English

Corpus description [READ](#)

**COUNTS**

Tokens: 1,043,519,660

**Words: 895,924,758**

**TRENDS**  🔍 ℹ️

**BASIC** **ADVANCED**

Structure attribute: Month

Subcorpus: none (the whole corpus)

Minimum frequency: 84

Maximum p-value: 0.01

Regular expression filter:

Attributes: word

Method: Mann-Kendall, Theil-Sen (all)

Sort items by: Trends

Filter out capitalized

Filter out non-words

No filter  
 Only positive trend  
 Only negative trend

	Month ↓	Frequency
1	<input type="checkbox"/>	267
2	<input type="checkbox"/>	===NONE
3	<input type="checkbox"/>	2014-04
4	<input type="checkbox"/>	2014-07
5	<input type="checkbox"/>	2016-11
6	<input type="checkbox"/>	2017-06
7	<input type="checkbox"/>	2017-11
8	<input type="checkbox"/>	2019-11
9	<input type="checkbox"/>	2020-01
10	<input type="checkbox"/>	2020-02
25	<input type="checkbox"/>	2022-01
26	<input type="checkbox"/>	2022-02
27	<input type="checkbox"/>	2022-03
28	<input type="checkbox"/>	2022-04
29	<input type="checkbox"/>	2022-05

There are a number of important differences between NOW (from English-Corpora.org) and Trends (from Sketch Engine):

Feature	NOW (English-Corpora.org)	Trends (Sketch Engine)
Size (as of May 2022)	15.1 billion words (~17x as large)	900 million words
Growth each month	7-8 million words	~4 million words
Corpus begins in	January <b>2010</b> (e.g. 244m words 2010, 305m in 2011)	~Jan <b>2020</b> (a few texts before then)
Search for	Words, <b>phrases, collocates</b>	Only single words
Granularity	Year, month, <b>and day</b>	Year and month

As mentioned, NOW is the only corpus that shows the frequency by year, word, and day. For example, the following are the results from a search for *Ukraine*, in the time period in which it was invaded by Russia (24 Feb 2022):

### Ukraine

[CHANGE TO VERTICAL CHART](#) / [CLICK TO SEE CONTEXT](#) [See frequency by country](#)

SECTION	ALL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
FREQ	918433	1937	2156	3801	3859	32279	15079	25781	27958	28273	70822	42979	48596	614913
WORDS (M)	15100	244.1	304.8	371.3	401.5	429.4	512.5	1,531.3	1,746.5	1,569.1	1,987.5	2,607.8	2,449.2	1,025.4
PER MIL	60.82	7.93	7.07	10.24	9.61	75.16	29.43	16.84	16.01	18.02	35.63	16.48	19.84	599.68
SEE ALL SUB-SECTIONS AT ONCE														

SECTION	22-01	22-02	22-03	22-04	22-05
FREQ	35040	153651	250149	134712	40949
WORDS (M)	272.6	242.1	229.0	206.8	78.4
PER MIL	128.56	634.65	1,092.50	651.57	522.31
CLICK FOR CONTEXT					

22-02-13	22-02-14	22-02-15	22-02-16	22-02-17	22-02-18	22-02-19	22-02-20	22-02-21	22-02-22	22-02-23	22-02-24	22-02-25	22-02-26	22-02-27	22-02-28
3556	4308	4705	4021	2112	5025	3850	2956	6782	8900	10698	22537	19955	11593	12934	10818
9.4	11.2	11.9	11.4	3.7	11.7	7.5	4.0	9.0	8.7	9.3	9.4	10.1	5.7	5.7	5.6
377.6	385.0	393.9	353.3	575.4	431.1	515.8	734.8	757.4	1,022.2	1,149.2	2,404.4	1,970.5	2,020.2	2,279.8	1,923.6

The ability to see results by day is important in finding the **earliest use of a word or phrase**. For example, the phrase *fake news* increases markedly in November 2016 – the date of the US elections. But did it increase before or after the elections on November 8? This is important, since people on both sides of the political spectrum use the word in different ways, and for different purposes. The data from NOW shows clearly that the normalized frequency (frequency per million words) is quite low every day before the elections, but that it increases seven-fold right after the elections (Nov 9 to Nov 10/11) and that it increases even more a week or so after the elections. NOW is the only corpus that allows us to search so much data in such a granular way.

fake news

CHANGE TO VERTICAL CHART / CLICK TO SEE CONTEXT See frequency by country

SECTION	ALL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
FREQ	128665	24	43	58	64	89	95	4894	25819	29502	25982	24392	12570	5133
WORDS (M)	15100	244.1	304.8	371.3	401.5	429.4	512.5	1,531.3	1,746.5	1,569.1	1,987.5	2,607.8	2,449.2	1,025.4
PER MIL	8.52	0.10	0.14	0.16	0.16	0.21	0.19	3.20	14.78	18.80	13.07	9.35	5.13	5.01

SECTION	16-01	16-02	16-03	16-04	16-05	16-06	16-07	16-08	16-09	16-10	16-11	16-12
FREQ	10	18	24	21	34	17	41	64	69	81	1697	2813
WORDS (M)	90.2	82.5	133.6	132.5	131.8	110.6	111.4	151.9	146.8	151.9	145.8	141.7
PER MIL	0.11	0.22	0.18	0.16	0.26	0.15	0.37	0.42	0.47	0.53	11.64	19.85

DAY	16-11-01	16-11-02	16-11-03	16-11-04	16-11-06	16-11-07	16-11-08	16-11-09	16-11-10	16-11-11	16-11-12	16-11-13	16-11-14	16-11-15	16-11-16	16-11-17	16-11-18
FREQ	4	2	2	10	6	12	7	8	56	61	17	59	130	188	71	102	69
WORDS (M)	5.4	5.5	5.3	5.3	4.5	4.8	5.5	5.2	5.2	6.0	5.0	4.5	4.7	4.3	5.0	5.2	5.8
PER MIL	0.7	0.4	0.4	1.9	1.3	2.5	1.3	1.5	10.8	10.3	3.4	13.2	27.7	43.5	14.2	19.7	12.0

Another important difference is that the Trends corpus from Sketch Engine only shows the frequency of words (and remember, just single words – not phrases). In NOW, on the other hand, we can also see the **collocates** of a given word or phrase (even at the level of individual months, weeks, or days). This provides important insight into *how* the word or phrase was being used at a particular point in time. For example, the following is a comparison of *climate NOUN* in 2020-2022 and 2010-2014. Notice the rise in phrases like *climate anxiety / breakdown / grief*:

SEC 1 (2022, 2019, 2020, 2021): 8,069,967,850 WORDS

SEC 2 (2010, 2011, 2012, 2013, 2014): 1,751,131,332 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	CLIMATE PROVISIONS	692	1	0.1	0.0	150.2	1	CLIMATE BLAME	37	3	0.0	0.0	56.8
2	CLIMATE ANXIETY	681	1	0.1	0.0	147.8	2	CLIMATE PARLIAMENT	34	11	0.0	0.0	14.2
3	CLIMATE EMERGENCY	16004	27	2.0	0.0	128.6	3	CLIMATE CAMP	77	26	0.0	0.0	13.6
4	CLIMATE BREAKDOWN	1521	3	0.2	0.0	110.0	4	CLIMATE EXCHANGE	62	27	0.0	0.0	10.6
5	CLIMATE SPENDING	472	1	0.1	0.0	102.4	5	CLIMATE DEBATE	970	730	0.6	0.1	6.1
6	CLIMATE PLEDGE	1843	4	0.2	0.0	100.0	6	CLIMATE SCEPTICS	170	178	0.1	0.0	4.4
7	CLIMATE COLLAPSE	301	1	0.0	0.0	65.3	7	CLIMATE COMMISSIONER	57	68	0.0	0.0	3.9
8	CLIMATE ACCORDS	491	2	0.1	0.0	53.3	8	CLIMATE DIPLOMACYBY	65	0	0.0	0.0	3.7
9	CLIMATE ACCORD	3892	17	0.5	0.0	49.7	9	CLIMATE FLUCTUATIONS	81	108	0.0	0.0	3.5
10	CLIMATE PLATFORM	218	1	0.0	0.0	47.3	10	CLIMATE SKEPTICS	112	159	0.1	0.0	3.2
11	CLIMATE PROTESTS	615	3	0.1	0.0	44.5	11	CLIMATE MODELERS	35	50	0.0	0.0	3.2
12	CLIMATE PROMISES	196	1	0.0	0.0	42.5	12	CLIMATE SCEPTIC	103	174	0.1	0.0	2.7
13	CLIMATE STRIKE	3427	0	0.4	0.0	42.5	13	CLIMATE SKEPTIC	65	113	0.0	0.0	2.7
14	CLIMATE EMERGENCIES	345	2	0.0	0.0	37.4	14	CLIMATE DEBT	45	79	0.0	0.0	2.6
15	CLIMATE GRIEF	168	1	0.0	0.0	36.5	15	CLIMATE TREATY	153	272	0.1	0.0	2.6

As mentioned, we can find and compare collocates at the level of individual months and even days. For example, the following are collocates of *fake news* in the month after the phrase increased in frequency (i.e. from 10 Nov 2016 through 10 Dec 2016). (And notice the misspelling of *ridiculous*, which comes from the "comments" sections of online newspapers, and which shows the sometimes informal nature of these comments).

List Chart **Collocates** Compare KWIC

fake news Word/phrase [POS] ?

\* Collocates [POS]

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

Find collocates Reset

[ - ] START 11/10/2016 to 12/10/2016 END

HELP	①	★	WORDS	FREQ
1	i	★	HOAXES	71
2	i	★	PROLIFERATION	44
3	i	★	PEDDLE	10
4	i	★	PURVEYORS	9
5	i	★	AMOK	8
6	i	★	POST-TRUTH	6
7	i	★	PROLIFERATED	6
8	i	★	HAND-WRINGING	5
9	i	★	PURVEYOR	5
10	i	★	REDICULOUS	4
11	i	★	ASSAIL	4
12	i	★	PIZZAGATE	4

The following are some of the concordance lines for this search (*fake news*, with the collocate *hoaxes*, from 16-11-10 to 16-12-10). Notice that even during this early period, the phrase had spread to many different countries, including Canada, the UK (GB), South Africa, India, and Malaysia.

1	16-12-07 US	WBUR	Q	saying, " We have already launched work enabling our community to flag <b>hoaxes</b> and <b>fake news</b> , and there is more we can do here. "
2	16-12-05 US	Engadget	Q	lead to violent acts. # The internet has always had its fair share of <b>fake news</b> and <b>hoaxes</b> , but it wasn't until recently that it's become an
3	16-12-04 US	Gizmodo	Q	one by that name working at the magazine. # In a year full of <b>fake news</b> and <b>hoaxes</b> , even a great white shark may seem like small potatoes.
4	16-12-04 MY	Malaysia Chronicle	Q	# Horner has been responsible for a variety of viral news <b>hoaxes</b> , including a <b>fake news</b> story that went viral about a lawsuit between reviews site Y
5	16-12-04 MY	Malaysia Chronicle	Q	from it, " said Paul Horner, who has made a living off viral <b>fake news hoaxes</b> , in an interview with The Washington Post. " Right now I
6	16-12-01 IN	Businessinsider India	Q	Africa, Patrick Walker. # Facebook says that it wants to get rid of <b>fake news</b> and <b>hoaxes</b> on its platform. But its plan to do so does not
7	16-11-29 AU	Gizmodo Australia	Q	of what people see on Facebook is authentic. Only a very small amount is <b>fake news</b> and <b>hoaxes</b> . The <b>hoaxes</b> that do exist are not limited to one pe
8	16-11-27 CA	Varsity	Q	the matter at one point -- stating that " only a very small amount is <b>fake news</b> and <b>hoaxes</b> " -- he and his company have since taken action on the
9	16-11-23 PH	Rappler	Q	revenues -- report # MANILA, Philippines -- In spite of the <b>hoaxes</b> and the <b>fake news</b> that have been hounding Facebook, the company has seen po
10	16-11-21 ZA	Media Update	Q	of what people see, is authentic. " Only a very small amount is <b>fake news</b> and <b>hoaxes</b> . " he wrote. # Corinne Podger, digital editorial capability
11	16-11-21 ZA	Media Update	Q	we see, because we have many more reasons to be. " # With <b>fake news</b> and <b>hoaxes</b> being shared across social media, news publishers can fill a un
12	16-11-20 GB	The Guardian	Q	than 99% of what people see is authentic. Only a very small amount is <b>fake news</b> and <b>hoaxes</b> . The <b>hoaxes</b> that do exist are not limited to one part

For even more granularity, we could create a "Virtual Corpus", based on metadata from the 26+ million texts in NOW (as of May 2022), including country, web domain (e.g. New York Times), title, date (down to the level of individual days), and words in the text. For example, we could create a Virtual Corpus of texts from the United States from Nov-Dec 2016, with the phrase *fake news* in the title of the article.

Web domain  Find sources  
(can use substring, e.g. Times, Houston)

Article title

Country   
United States  
Canada  
Great Britain  
Ireland

Dates  to

Words in text

About 1-2 seconds later, we have a list of the best texts. We simply provide a name, e.g. *fake\_news*.

SAVE AS: fake\_news OR ADD TO: --SELECT-- SUBMIT RESET

HELP	<input type="checkbox"/>	98	DATE	COUNTRY	WEBSITE	TITLE
1	<input checked="" type="checkbox"/>		16-11-11	United States	CBS News	With Trump's election, Facebook wrestles with the ....
2	<input checked="" type="checkbox"/>		16-11-14	United States	TechCrunch	Dear Facebook, please fix the plague of fake news ....
3	<input checked="" type="checkbox"/>		16-11-14	United States	The Verge	Google will soon ban fake news sites from using it....
4	<input checked="" type="checkbox"/>		16-11-15	United States	Huffington Post	Facebook, Google Take Small Steps To Stop Spread O....
5	<input checked="" type="checkbox"/>		16-11-15	United States	The Daily Dot	3 Chrome extensions to cut fake news out of your l....
6	<input checked="" type="checkbox"/>		16-11-16	United States	9NEWS.com	Breitbart, Occupy Democrats among list of alleged ....
7	<input checked="" type="checkbox"/>		16-11-16	United States	KTLA	To Keep Fake News Out of Facebook Newsfeed, Colle...

Once we have created the Virtual Corpus, we can search within the Virtual Corpus (for words, phrases, collocates, etc), and we can compare the frequency of words or phrases in different Virtual Corpora (for example, in the US and the UK in the same period). In addition, in just 1-2 seconds more, we can generate a list of keywords from the Virtual Corpus:

FAKE\_NEWS [76,841 WORDS, 98 TEXTS] (0.5% OF TOTAL) **NOUN** VERB ADJ ADV N+N ADJ+N [ALL CORPORA] SAVE LIST

HELP	SAVE	WORD (CLICK FOR CONTEXT) TRANSLATE ALL ENTRIES	FREQ	# TEXTS	SPECIFIC		ENTIRE CORPUS	EXPECTED
					FREQ	TEXTS		
					22.5	9		
1	★	HOAX	37	19	0.9		7,807	39.7
2	★	CONSPIRACY	65	26	0.3		50,239	255.7
3	★	LIE	40	16	0.2		52,193	265.6
4	★	PROPAGANDA	24	13	0.1		36,617	186.3
5	★	PIZZA	27	19	0.1		42,522	216.4
6	★	RIFLE	23	16	0.1		37,601	191.3
7	★	NEWS	734	81	0.1		1,223,971	6,228.6
8	★	SPREAD	23	15	0.1		44,127	224.6

And then when we click on one of these keywords (e.g. *hoax*), we see results just from the Virtual Corpus that we have created:

1	16-12-05 US	Chicago Sun-Times	🔍	🔍	🔍	Sunday, Dec. 4, 2016. A man who said he was investigating a <b>conspiracy</b> theory about Hillary Clinton running a child sex ring out of the pizza place fired
2	16-12-21 US	WIRED	🔍	🔍	🔍	lined up in Austin became proof that Democrats were bringing protestors to Trump rallies. <b>Conspiracy</b> theorists say a screenshot from a video of Preside
3	16-12-09 US	Salon	🔍	🔍	🔍	, liberals and leftists do this, too -- as anyone cornered by an anti-vaccine <b>conspiracy</b> theorist can tell you -- but the problem, at least in recent years,
4	16-12-05 US	Winston-Salem Journal	🔍	🔍	🔍	with a dangerous weapon, told police he was there to " self-investigate " a <b>conspiracy</b> which likely involved internet rumors that Hillary Clinton's presiden
5	16-12-21 US	Phys.Org	🔍	🔍	🔍	Americans do, you are exposed to a daily dose of hoaxes, rumors, <b>conspiracy</b> theories and misleading news. When it's all mixed in with reliable informati
6	16-12-05 US	Chicago Sun-Times	🔍	🔍	🔍	could lead to violence. " # The nasty web of fake news, sinister <b>conspiracy</b> theories, online trolls and outright lies has been building for a long time.
7	16-12-06 US	Salon	🔍	🔍	🔍	WikiLeaked emails. # Welch told police he'd planned to " self-investigate " the <b>conspiracy</b> theory and, according to court documents, surrendered when f
8	16-12-06 US	CBS News	🔍	🔍	🔍	# Maybe I'm wrong about that, but I doubt it: Trump embraced <b>conspiracy</b> theorists like Alex Jones from the start of his campaign. His longtime political c
9	16-12-05 US	The Hill	🔍	🔍	🔍	in a statement to The Washington Post the arrest shows " promoting false and reckless <b>conspiracy</b> theories comes with consequences. " # " I hope that tl
10	16-12-27 US	Media Matters for Americ	🔍	🔍	🔍	working on the Trump campaign to share links from sites that push fake news and <b>conspiracy</b> theories. He was also responsible for an anti-Semitic Trum
11	16-12-06 US	CBS News	🔍	🔍	🔍	the extreme right wing movement that favors white supremacy and the place where these new <b>conspiracy</b> theories are now taking root, doesn't have su

Finally, as we discuss in [another help file](#), in NOW we can quickly and easily find the keywords for individual days. For example, the following are the keywords for 25 Feb 2022, the day after Ukraine was invaded.

RESULTS FOR 22-02-26 (compared: 22-02-10 to 22-02-12) 22-05-10 22-05-09 22-05-08 22-05-07 Include capitalized words

	Word (click to see)	PoS	Freq	Texts	Countries	Ratio
1	missiles	NOUN	345	199	14	595.58
2	explosions	NOUN	216	158	10	372.88
3	invaded	VERB	205	190	8	353.89
4	airspace	NOUN	202	73	6	348.72
5	flee	VERB	174	141	10	300.38
6	evacuate	VERB	161	116	10	277.94
7	fleeing	VERB	160	107	5	276.21
8	unprovoked	ADJ	150	118	8	258.95

(The keywords – along with all of the texts – are created and processed by 10 PM. In this sense, the corpus is never more than just a few *hours* out of date.)

We should also mention that if you are interested primarily in data related to COVID / Coronavirus, you may want to consider the **Coronavirus Corpus**. It contains about 1.5 billion words in about 2 million texts (through May 2022), and it is growing by about 30-50 million words per month. All of the functionality shown above for the NOW Corpus is also available in the Coronavirus Corpus. (See [Davies \(2021\)](#) for an overview of the corpus in the *International Journal of Corpus Linguistics*).

In summary, no other corpus serves as such a powerful monitor corpus of English as the NOW Corpus. It is much larger than comparable corpora, it provides a much wider range of searches, and it provides much more granularity than any other corpus. And it is freely available (along with many other very useful corpora of English) from English-Corpora.org.