

Artículo de Investigación

Extracción de palabras clave y expresiones multipalabra de un corpus textual de noticias sobre cambio climático

Keyword and multi-word terms extraction from a textual corpus of news about climate change

Vanesa Álvarez Torres¹: Universidad de Cádiz, España.

vanesa.alvarez@uca.es

Fecha de Recepción: 05/06/2024

Fecha de Aceptación: 13/08/2024

Fecha de Publicación: 29/10/2024

Cómo citar el artículo:

Álvarez Torres, V. (2024). Extracción de las palabras clave y expresiones multipalabra de un corpus textual de noticias sobre cambio climático [Keyword and multi-word terms extraction from a textual corpus of news about climate change]. *European Public & Social Innovation Review*, 9, 1-17. <https://doi.org/10.31637/epsir-2024-1133>

Resumen:

Introducción: En este trabajo presentamos la confección de un corpus textual sobre cambio climático y la extracción de las palabras clave y las expresiones multipalabra para su análisis.

Metodología: Tanto para la elaboración del corpus textual como para la extracción de las listas de palabras clave y expresiones multipalabras hemos usado Sketch Engine. Hemos recopilado los textos que componen el corpus de la web y, posteriormente, nos hemos descargado las listas de palabras clave. **Resultados y análisis:** Los datos resultantes muestran que, además de las palabras más frecuentes que se suelen usar al hablar de cambio climático (como *calentamiento* y *calor*), una gran parte de las palabras clave son nombres de personas que, desde la política, la ciencia o el activismo, han tratado de un modo u otro, este tema. Por tanto, depende del momento social en el que se publiquen los textos. Por el contrario, las expresiones multipalabra no contienen nombres propios, están formadas por elementos léxicos.

Conclusiones: A partir del trabajo realizado nos hemos acercado a un tema tan actual y transversal como es el cambio climático. Pretendemos aprovechar este corpus en futuros estudios.

Palabras clave: corpus textual; noticias; cambio climático; extracción; palabras clave; expresiones multipalabra.

¹ Autor Correspondiente: Vanesa Álvarez Torres. Universidad de Cádiz (España).

Abstract:

Introduction: In this paper we present the elaboration of a textual corpus on climate change and the extraction of keywords and multi-word terms to analyze it. **Methodology:** We used Sketch Engine both for the elaboration of the text corpus and for its extraction. We have compiled the texts from the web and, subsequently, we have downloaded the lists of keywords. **Results and analysis:** The resulting data demonstrate that, in addition to the most frequent words used when talking about climate change (like “calentamiento” and “calor”), a large part of the keywords are names of people who, from politics, science or activism, have dealt with this issue in one way or another. Therefore, it depends on the social moment in which the texts are published. On the other hand, multi-word terms do not contain proper names, they are made up of lexical elements. **Conclusions:** We have approached the very topical and cross-curricular issue of climate change based on the work carried out. We intend to use this corpus in future studies.

Keywords: text corpus; news; climate change; extraction; keywords; multi-word terms.

1. Introducción

La concienciación por la lucha contra el cambio climático constituye un tema social de interés que alcanza, podríamos atrevernos a decir, todas las esferas de la sociedad, de ahí su transversalidad. Como se define en el *Glosario sobre cambio climático para el aula*, el cambio climático “es un fenómeno complejo con una terminología que abarca diferentes campos del conocimiento: química, biología, ciencias ambientales, economía, sociedad, filosofía, historia...” (UNICEF, 2017, p. 3).

En la actualidad, el impulso por la educación contra el cambio climático viene avalado por la labor que realiza las Naciones Unidas con el establecimiento de los Objetivos de Desarrollo Sostenible, los denominados ODS, en la Agenda 2030. En relación con el tema que nos ocupa en este trabajo, el objetivo 13, “acción por el clima”, es el que se vincula con el cambio climático. Este organismo internacional cuenta con la denominada, en español, Convención Marco de las Naciones Unidas sobre el Cambio Climático, United Nations Framework on Climate Change (UNFCCC), en inglés, cuyo origen tiene lugar a partir del Protocolo de Kioto, con fecha de aprobación el 11 de diciembre de 1997 y entrada en vigor unos años más, el 16 de febrero de 2005 (United Nations Framework on Climate Change, s. f., en línea). Con el fin de continuar con los compromisos de 2013 a 2020, se añadió la Enmienda de Doha a este protocolo y fue aprobada el 8 de diciembre de 2012 (United Nations Framework on Climate Change, s. f., en línea).

Asimismo, la Conferencia de las Partes de la Convención Marco de las Naciones Unidas sobre el Cambio Climático (COP) aboga por el cambio climático desde 1995, que tuvo lugar la primera COP en Berlín (Alemania). La COP28 es la última que ha tenido lugar, se celebró en Dubai (Emiratos Árabes Unidos) del 30 de noviembre al 12 de diciembre de 2023 (United Nations Framework on Climate Change, s. f., en línea). Además de la COP, otros encuentros entre expertos y líderes internacionales que tratan de apostar por el medio ambiente son las Convenciones de Río y la Cumbre de la Tierra.

En lo que respecta a los corpus textuales, constituyen un recurso documental que permiten al investigador extraer información de una manera más ágil, cuantitativa y objetiva, para, posteriormente, realizar un análisis de los datos obtenidos. En este trabajo nos aproximamos al estudio de este tema transversal que incumbe a toda la sociedad, el cambio climático, a partir de un corpus de textos periodísticos. En esta línea existe una variedad de autores que se han centrado en el análisis del discurso del cambio climático a partir de un corpus de textos

periodísticos, como, por ejemplo, Grundmann y Ramesh (2010, pp. 113-113), que realizan un estudio a partir de un corpus de noticias de Estados Unidos, Reino Unido, Francia y Alemania descargadas de Nexis para comparar el tratamiento del cambio climático y la influencia de las políticas sobre cambio climático vigentes en cada uno de estos países. Otro trabajo más actual es el de Gillings y Darrell (2023, pp. 111-133), que investigan el uso de las expresiones “*climate change*” y “*global warming*” en la prensa británica de 2003 a 2019. Por su parte, Liu y Huang (2022, pp. 177-188) también estudian “*climate change*” y “*global warming*”, pero en el periódico estadounidense *The New York Times*. También existen publicaciones sobre la cobertura de este tema a nivel internacional² y de manera comparativa, como en Barkemeyer *et al.* (2017, pp. 1029-1054), o sobre la construcción de un corpus específico sobre esta temática. En esta última línea de investigación señalada, Volkanovska *et al.* (2023, pp. 177-188) explican el proceso de construcción de un corpus piloto multimodal para the InsightsNet Climate Change Corpus (ICCC) haciendo uso del procesamiento del lenguaje natural (PLN). Además, no solo existen estudios con corpus de textos periodísticos, también pertenecientes a otros ámbitos, como, por ejemplo, la política. Tal es el caso del trabajo de Willis (2017, pp. 212-231), que analiza cómo hablan los políticos sobre cambio climático, concretamente, toma la ley sobre cambio climático de Reino Unido *The Climate Change Act 2008*.

Desde nuestro ámbito de investigación, la lingüística, nos hemos propuesto como objetivo en este trabajo extraer las palabras más usadas a partir de un análisis de la frecuencia de aparición en noticias en formato digital sobre cambio climático. Nuestro interés no radica en la comparación entre los medios donde aparecen publicadas las noticias, sino en las listas extraídas del corpus compilado. Así pues, nos hemos planteado la búsqueda de las palabras más frecuentes para, en futuros estudios, realizar una comparación con el léxico disponible del centro de interés “cambio climático” de jóvenes de la provincia de Cádiz, recopilado a partir de pruebas de disponibilidad léxica en el marco del proyecto de investigación “Observación del Pulso Social en Andalucía a través del Análisis Léxico (Pulso Andaluz). Primera Fase (UMA20-FEDERJA-013)”, en el que hemos participado. Por motivos de extensión, nos limitamos en este primer estudio sobre el cambio climático al corpus textual de noticias en determinados periódicos y canales de televisión y radiotelevisión, que será de gran utilidad para su comparación con los resultados de otras investigaciones de carácter sociolingüístico, como hemos mencionado anteriormente.

En los siguientes apartados detallaremos cada uno de los pasos llevados a cabo, así como también los resultados, el análisis y las conclusiones obtenidas con este estudio.

2. Metodología

En la elaboración y el estudio de este corpus hemos seguido la misma metodología que en Sánchez-Saus Laserna y Álvarez Torres (2022, pp. 73-98). Si bien el tipo y la temática de los textos eran diferentes, pues analizamos artículos de opinión que divulgan sobre cuestiones lingüísticas, la metodología empleada en la recopilación de textos planos y la extracción de las listas de palabras clave puede aplicarse a otros tipos de textos con otras temáticas. En el presente trabajo hemos creado carpetas dentro del propio corpus para cada uno de los enlaces. El corpus que hemos confeccionado para este estudio está compuesto por 405.463 palabras aproximadamente. En concreto, las fuentes en las que aparecen los textos escritos de noticias sobre cambio climático han sido publicadas en páginas web de periódicos en versión digital y canales de televisión y radiotelevisión del 24 de septiembre de 2023 al 24 de octubre de 2023, el último día coincide con el Día Internacional contra el Cambio Climático. En un principio,

² De manera más general, se encuentra el informe realizado por Boykoff y Roberts (2007) para las Naciones Unidas, titulado “Media coverage of climate change: Current trends, strengths, weaknesses”.

empezamos recopilando solo los textos publicados el 24 de octubre de 2023 con la finalidad de explorar de qué se habla en ese día especial para la lucha contra el cambio climático, pero al observar que no eran suficientes fuentes, ampliamos el rango temporal a un mes. Debido a la laboriosa tarea de recuperar manualmente cada uno de los enlaces para la extracción de los textos en .txt, optamos por limitar la extracción a un mes para este primer trabajo. No obstante, no se trata de un corpus cerrado, podemos seguir ampliándolo.

Todos los textos que componen el corpus están escritos en español peninsular y el ámbito de los medios en los que aparecen publicados es nacional (España) y, en menor número de casos, regional (Andalucía) y local (Cádiz). Cada una de estas fuentes cuenta con un enlace, que hemos introducido en la aplicación de creación de corpus de la plataforma Sketch Engine para confeccionar un corpus introduciendo las URL y agrupando los textos planos en carpetas, clasificadas por marcas de medio de comunicación, esto es, una carpeta³ por cada marca, de tal manera que contamos con 16 carpetas en total.

El número total de enlaces introducidos para extraer las fuentes ha sido 557, no vamos a añadir cada uno de los enlaces que nos lleva a la fuente por motivos de extensión. No obstante, incluimos a continuación los enlaces generales de los canales o periódicos digitales en los que se han publicado los textos escritos sobre cambio climático:

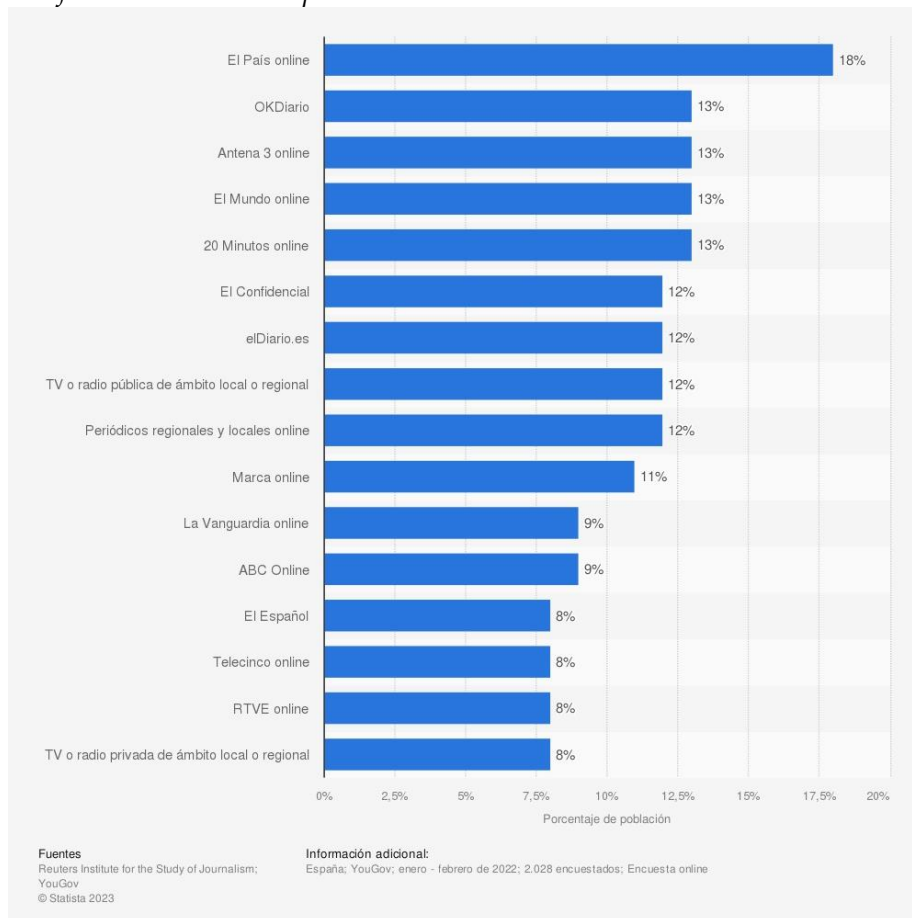
1. El País online: <https://elpais.com/noticias/cambio-climatico>
2. OKDiario: <https://okdiario.com/noticias/cambio-climatico/>
3. Antena 3 online:
https://www.antena3.com/noticias/temas/cambio_climatico-1
4. El Mundo online: <https://www.elmundo.es/t/ca/cambio-climatico.html>
5. 20Minutos online: <https://www.20minutos.es/minuteca/cambio-climatico/1/>
6. El confidencial: <https://www.elconfidencial.com/tags/temas/cambio-climatico-6047/>
7. elDiario.es: <https://www.eldiario.es/temas/cambio-climatico/>
8. Marca online: <https://www.marca.com/>
9. La Vanguardia online: <https://www.lavanguardia.com/topics/cambio-climatico>
10. ABC online: <https://www.abc.es/natural/cambio-climatico/>
11. El Español: https://www.elespanol.com/temas/cambio_climatico/
12. Telecinco online: <https://www.telecinco.es/tags/cambio-climatico/>
13. RTVE online: <https://www.rtve.es/temas/cambio-climatico/1081/>
14. RTVA online: <https://www.canalsur.es/noticias/medioambiente-4373.html>
15. Diario de Cádiz online:
https://www.diariodecadiz.es/tag/cambio_climatico/
16. La Voz de Cádiz online: <https://www.lavozdigital.es/natural/cambio-climatico/>

A la hora de seleccionar las marcas de periódicos y canales de televisión y radiotelevisión, hemos consultado el “ranking de las principales marcas de medios de comunicación online según el porcentaje de población que las usaba de forma semanal en España en 2022” realizado por Statista (2023). Recogemos a continuación la infografía con las marcas y los porcentajes:

³ Los metadatos de cada carpeta han consistido en el nombre del canal de televisión o radiotelevisión o del periódico digital.

Figura 1.

Ranking de las principales marcas de medios de comunicación online según el porcentaje de población que las usaba de forma semanal en España en 2022.



Fuente: [Statista Search Department \(2023\)](#)

Hemos tratado de adaptar estos datos a nuestro entorno regional y local con el fin de ajustar el presente estudio a nuestro ámbito geográfico. El guiarnos por las marcas más usadas por los ciudadanos, y su consecuente adecuación a nuestro entorno, ha conllevado a que este corpus textual sea heterogéneo.

En lo que respecta a la labor de extracción de textos para el corpus, en esta tabla recogemos el número de enlaces introducidos por cada marca del medio de comunicación y el número total de palabras que suman los textos extraídos:

Tabla 1.

Nº de palabras y nº de enlaces por marca de medio de comunicación.

	Marca del medio de comunicación	Nº enlaces	Nº Palabras
1	El País online	167	~164.652
2	OKDiario	11	~7.984
3	Antena 3 online	11	~6.080
4	El Mundo online	17	~8.706
5	20Minutos online	18	~10.750
6	El Confidencial	31	~16.580
7	elDiario.es	74	~82.928

8	Marca online	1	~338
9	La Vanguardia	26	~27.555
10	ABC online	75	~1.478
11	El Español	42	~35.598
12	Telecinco online	14	~8.176
13	RTVE online	56	~28.702
14	RTVA online	4	~616
15	Diario de Cádiz online	6	~3.549
16	La Voz de Cádiz online	4	~1.771
TOTAL		557	~405.463

Fuente: Elaboración propia (2024).

Una vez creado el corpus, hemos extraído las palabras clave, tanto las palabras únicas y las combinaciones léxicas (*multi-word terms*) también con Sketch Engine. Para ello, hemos usado un corpus de referencia, el Spanish Web 2018 (esTenTen18), de este modo, las palabras clave extraídas resultan ser más frecuentes en nuestro corpus (*focus*) que en el de referencia que en el corpus de comparación (*reference*), teniendo en cuenta el *keyness score* (Kilgarrif *et al.*, 2014). La fórmula para calcular el *keyness score*, tal y como señalan en Sketch Engine, a partir de Kilgarriff (2009), es la siguiente:

$$\frac{fpm_{rmfocus} + N}{fpm_{rmref} + N}$$

where

$fpm_{rmfocus}$ is the normalized (per million) frequency of the word in the focus corpus,

fpm_{rmref} is the normalized (per million) frequency of the word in the reference corpus,

N is the smoothing parameter ($N = 1$ is the default value)⁴.

En cuanto al corpus de referencia, el Spanish Web 2018 (esTenTen18), está compuesto por textos de internet de páginas web en español de España y América, que suman un total de 16,9 billones de palabras, y forma parte de TenTen Corpus Family⁵.

Una vez descargadas las palabras clave, editamos las listas de palabras y eliminamos aquellas palabras o expresiones repetitivas que no tienen contenido léxico, ya que no resulta relevante para nuestro estudio, como pueden ser las marcas de canales o periódicos. En el caso de los nombres propios (por ejemplo, *Copernicus*) y nombres de empresa (por ejemplo, *TotalEnergies*), no los hemos eliminado del corpus, puesto que pueden resultar de interés para otros estudios en los que vayamos más allá de las palabras con contenido léxico. El número total de palabras clave recogidas en las listas extraídas es de 15.225 y el de expresiones multipalabra 46.383.

Una vez descargadas las listas de palabras, procedimos a la edición de los ítems, que, tenemos que añadir que solo ha necesitado correcciones la lista de palabras únicas, las expresiones multipalabra no. Las modificaciones realizadas son las siguientes:

- Mayúscula inicial: Copernicus, Dubái, Lancet, Sunak, Romanello, TotalEnergies, Codorníu, Guterres, Birol, Antártida, Thunberg, Ederer Mazón, Pnuma, Amazonia, Espuña, Burgess, Ártico, Doñana, Greta, Jaber, Milei, Vox, Agostinho, Banga,

⁴ <https://www.sketchengine.eu/documentation/simple-maths/>

⁵ <https://www.sketchengine.eu/estenten-spanish-corpus/>

Emiratos⁶, Greenpeace, Scholz, Faith, Groenlandia, António, Naturgy, Soina, Taalas y Pradas.

- Mayúsculas: COP28, ZBE, AEMET, OMM, IPCC, C3S, CO2, AIE, TEDH, PATIVEL, OC, COP y PNIEC.
- Singular: emisión, gigatonelada y ecologista.

Hemos usado una hoja de cálculo de Excel para realizar los cálculos y obtener los datos cuantitativos sobre los ítems que integran las dos listas extraídas.

3. Resultados y análisis

En el siguiente apartado recogemos dos listas (palabras clave y combinaciones léxicas) con los 100 primeros ítems, ya que la lista de palabras clave cuenta con un total de 15.225 ítems y la de expresiones multipalabra con 46.383 ítems, un número elevado para adecuarnos a los límites de extensión de este trabajo. Presentamos a continuación una lista con las 100 primeras palabras clave ordenadas de mayor a menor *score*:

Tabla 2.

100 primeras palabras clave (únicas).

	Ítem	Frecuencia (focus)	Frecuencia (reference)	Frecuencia relativa (focus)	Frecuencia relativa (reference)	Score
1	COP28	122	0	262,0455	0	263,05
2	preindustrial	81	5723	173,98103	0,29209	135,43
3	climático	2083	633883	4474,105	32,35238	134,18
4	Copernicus	53	1387	113,83944	0,07079	107,25
5	ZBE	45	80	96,65613	0,00408	97,259
6	AEMET	76	15874	163,24146	0,81018	90,732
7	descarbonización	44	2720	94,50822	0,13882	83,866
8	lúpulo	68	15519	146,05815	0,79206	82,061
9	calentamiento	346	178357	743,17822	9,10306	73,659
10	descarbonizar	35	1465	75,17699	0,07477	70,877
11	fósil	291	169050	625,04297	8,62804	65,023
12	Dubái	55	19741	118,13527	1,00755	59,344
13	Lancet	41	10540	88,06447	0,53794	57,911
14	Sunak	26	9	55,84576	0,00046	56,82
15	sequía	268	179659	575,64093	9,16951	56,703
16	pandemia	71	35149	152,50189	1,79395	54,941
17	Romanello	25	302	53,69785	0,01541	53,868
18	borrasca	41	13181	88,06447	0,67274	53,245
19	TotalEnergies	24	0	51,54993	0	52,55
20	Codorníu	27	2597	57,99368	0,13255	52,089
21	Guterres	32	7303	68,73325	0,37273	50,799
22	emisión	35	12112	75,17699	0,61818	47,076
23	invernadero	183	145878	393,06824	7,44538	46,661
24	OMM	26	4998	55,84576	0,25509	45,292

⁶ No hemos cambiado *Emiratos* al singular porque se trata de un nombre propio.

25	countdown	24	3399	51,54993	0,17348	44,781
26	Birol	20	563	42,95828	0,02873	42,73
27	climate	32	12441	68,73325	0,63497	42,651
28	covid	19	3	40,81036	0,00015	41,804
29	derretimiento	25	6389	53,69785	0,32608	41,248
30	Antártida	76	60559	163,24146	3,09083	40,149
31	Thunberg	18	323	38,66245	0,01649	39,019
32	eucalipto	52	37536	111,69153	1,91578	38,649
33	Ederer	17	90	36,51454	0,00459	37,343
34	Mazón	20	4336	42,95828	0,2213	35,993
35	IPCC	30	16643	64,43742	0,84943	35,382
36	percebe	22	7145	47,25411	0,36467	35,36
37	C3S	16	103	34,36662	0,00526	35,182
38	CO2	161	174728	345,81415	8,91784	34,969
39	renovable	310	354632	665,85333	18,09985	34,914
40	AIE	26	12310	55,84576	0,62828	34,911
41	Pnuma	23	8711	49,40202	0,4446	34,89
42	Amazonia	39	28967	83,76865	1,47843	34,203
43	posidonia	20	5851	42,95828	0,29863	33,85
44	España	18	3604	38,66245	0,18394	33,5
45	Burgess	19	5227	40,81036	0,26678	33,005
46	meteorológico	145	169964	311,44751	8,67469	32,295
47	gigatonelada	15	1013	32,21871	0,0517	31,586
48	negacionista	19	6501	40,81036	0,3318	31,394
49	deshielo	32	24231	68,73325	1,23671	31,177
50	Ártico	57	58568	122,43109	2,98922	30,941
51	cálido	239	306243	513,35144	15,63015	30,929
52	Doñana	33	26095	70,88116	1,33185	30,826
53	ecologista	33	26988	70,88116	1,37742	30,235
54	Greta	21	10461	45,10619	0,53391	30,058
55	resiliencia	45	44748	96,65613	2,28387	29,738
56	dióxido	69	81732	148,20605	4,17147	28,852
57	Jaber	14	1713	30,0708	0,08743	28,573
58	inhabitable	19	9154	40,81036	0,46721	28,497
59	Milei	13	429	27,92288	0,0219	28,303
60	veintisiete	16	5053	34,36662	0,2579	28,116
61	glaciar	85	108807	182,57269	5,55334	28,012
62	TEDH	18	8818	38,66245	0,45006	27,352
63	negacionismo	15	4278	32,21871	0,21834	27,265
64	ecologismo	20	12072	42,95828	0,61614	27,2
65	Vox	32	30765	68,73325	1,5702	27,131
66	deforestación	46	52820	98,80404	2,69585	27,004
67	Agostinho	13	1610	27,92288	0,08217	26,727
68	calor	592	925108	1271,5651	47,21604	26,393
69	PATIVEL	12	515	25,77497	0,02628	26,089
70	quebrantahuesos	16	7055	34,36662	0,36008	26,003
71	Banga	12	591	25,77497	0,03016	25,991

72	mitigación	46	56432	98,80404	2,8802	25,721
73	hielo	249	395010	534,83057	20,16068	25,322
74	caluroso	66	91167	141,76231	4,65302	25,254
75	OC	54	71794	115,98735	3,66425	25,082
76	Emiratos	37	44455	79,47282	2,26891	24,618
77	temperatura	829	1402253	1780,6206	71,56876	24,551
78	torrencial	26	26104	55,84576	1,33231	24,373
79	Greenpeace	40	50443	85,91656	2,57453	24,316
80	Scholz	12	2062	25,77497	0,10524	24,225
81	Fatih	12	2123	25,77497	0,10835	24,157
82	COP	28	31626	60,14159	1,61414	23,389
83	carbono	214	370537	459,65359	18,91162	23,135
84	Groenlandia	23	23144	49,40202	1,18123	23,107
85	meteorólogo	20	17854	42,95828	0,91124	23
86	ecocidio	12	3266	25,77497	0,16669	22,949
87	António	15	8823	32,21871	0,45031	22,905
88	Naturgy	10	0	21,47914	0	22,479
89	PNIEC	10	1	21,47914	0,00005	22,478
90	Soina	10	6	21,47914	0,00031	22,472
91	supercontinente	11	2005	23,62705	0,10233	22,341
92	Taalas	10	165	21,47914	0,00842	22,291
93	geotermia	14	7957	30,0708	0,40611	22,097
94	récord	213	387147	457,50568	19,75937	22,087
95	precipitación	102	180432	219,08722	9,20896	21,558
96	meteorología	39	57940	83,76865	2,95717	21,422
97	Pradas	11	2943	23,62705	0,15021	21,411
98	combustible	342	656513	734,58655	33,50737	21,317
99	metano	28	36994	60,14159	1,88811	21,17
100	ola	204	391303	438,17444	19,97148	20,942

Fuente: Elaboración propia a partir de la herramienta de extracción de palabras clave de Sketch Engine (2024).

Si clasificamos estos 100 primeros ítems por clases de palabras, estarían agrupados de la siguiente manera, ordenados alfabéticamente:

Tabla 3.

Clasificación de las palabras clave.

Nombres propios	Nombres comunes	Adjetivos	Verbos
AEMET	borrasca	cálido	descarbonizar
Agostinho	calentamiento	caluroso	
AIE	calor	climático	
Amazonia	carbono	ecologista	
Antártida	climate	inhabitable	
António	CO2	meteorológico	
Ártico	combustible	negacionista	
Banga	countdown	preindustrial	

Birol	covid	renovable
Burgess	deforestación	torrencial
C3S	derretimiento	
Codornú	descarbonización	
COP	deshielo	
COP28	dióxido	
Copernicus	ecocidio	
Doñana	ecologismo	
Dubái	emisión	
Ederer	eucalipto	
Emiratos	fósil	
España	geotermia	
Fatih	gigatonelada	
Greenpeace	glaciar	
Greta	hielo	
Groenlandia	invernadero	
Guterres	lúpulo	
IPCC	metano	
Jaber	meteorología	
Lancet	meteorólogo	
Mazón	mitigación	
Milei	negacionismo	
Naturgy	OC	
OMM	ola	
PATIVEL	pandemia	
PNIEC	percebe	
Pnuma	posidonia	
Pradas	precipitación	
Romanello	quebrantahuesos	
Scholz	récord	
Soina	resiliencia	
Sunak	sequía	
Taalas	supercontinente	
TEDH	temperatura	
Thunberg	veintisiete	
TotalEnergies	ZBE	
Vox		

Fuente: Elaboración propia (2024).

Hemos conservado los nombres propios en la lista de palabras clave por si son de interés en futuros estudios. Además, si bien este tipo de palabras carece de contenido léxico, nos encontramos con nombres (y/o apellidos) de personas como políticos, activistas e investigadores, entre otros, que nos proporcionan información sobre las distintas esferas de la sociedad y se relacionan, de alguna manera, con el cambio climático. Los ítems que recogemos bajo este tipo son los siguientes: *Agostinho, António, Banga, Birol, Burgess, Ederer, Fatih, Greta, Guterres, Jaber, Mazón, Milei, Pnuma, Pradas, Romanello, Scholz, Soina, Sunak, Taalas y Thunberg*⁷.

⁷ Si bien los ítems *Greta* y *Thunberg* hacen referencia a una misma persona, hemos conservado ambas formas en el corpus confeccionado.

Otro grupo que hemos distinguido dentro de los nombres hace referencia a agencias, planes, grupos de expertos, conferencias, etc., como son: AEMET, AIE, C3S, COP, COP28⁸, Copernicus, IPCC, Lancet, OMM, PATIVEL, PNIEC y TEDH. También nos encontramos con nombres propios de lugares (Amazonia, Antártida, Ártico, Doñana, Dubái, Emiratos, España y Groenlandia), empresas (Codorníu, Naturgy y TotalEnergies), ONG (Greenpeace) y partidos políticos (Vox).

Todos estos nombres propios nos proporcionan información sobre distintas cuestiones que atañen a la sociedad, como la política o la economía, por ejemplo. Visto de este modo, se trata de un reflejo de la sociedad, concretamente, en el momento en el que se publicaron los textos extraídos, esto es, del 24 de septiembre de 2023 al 24 de octubre de 2023.

Antes de añadir la lista de expresiones multipalabra, creemos oportuno indicar que algunas de las palabras clave constituyen un elemento léxico que, combinado con otro/s, forma una expresión multipalabra (*multi-word term*), de ahí que algunas de ellas puedan aparecer también en la lista de combinaciones léxicas. Tal es el caso de *climático/a*, que, al unirse a *crisis* y *cambio*, componen las expresiones multipalabra *crisis climática* y *cambio climático*.

Añadimos a continuación la lista de combinaciones léxicas ordenadas de mayor a menor *score*, recordamos que se trata de una versión reducida, puesto que solo aparecen los 100 primeros ítems:

Tabla 4.

100 primeras combinaciones léxicas.

	Ítem	Frecuencia (focus)	Frecuencia (reference)	Frecuencia relativa (focus)	Frecuencia relativa (reference)	Score
1	crisis climática	161	3632	345,81415	0,18537	292,578
2	ola de calor	181	17549	388,77243	0,89567	205,612
3	combustible fósil	266	49411	571,34509	2,52186	162,512
4	cambio climático	1109	324417	2382,03638	16,55773	135,726
5	acción climática	59	2169	126,72692	0,11107	114,997
6	hielo marino	60	4498	128,87483	0,22957	105,626
7	gas de efecto invernadero	146	45697	313,59543	2,3323	94,408
8	emergencia climática	45	1288	96,65613	0,06574	91,632
9	calentamiento global	177	64545	380,18076	3,29427	88,765
10	transición energética	53	6014	113,83944	0,30694	87,869
11	lucha contra el cambio climático	66	13141	141,76231	0,6707	85,451
12	temperatura global	51	5934	109,54361	0,30286	84,847
13	calor extremo	47	4060	100,95196	0,20722	84,452
14	aumento de la temperatura	62	12183	133,17067	0,6218	82,729
15	nivel preindustrial	39	1033	83,76865	0,05272	80,523
16	efecto invernadero	168	68666	360,84955	3,5046	80,329
17	efecto del cambio climático	64	15739	137,46649	0,80329	76,785

⁸ No hemos unificado COP y COP28, puesto que la segunda de ella es más específica, indica el número de la edición de la Conferencia de las Partes de la Convención Marco de las Naciones Unidas sobre el Cambio Climático.

18	herramienta contra el cambio climático	32	21	68,73325	0,00107	69,659
19	cumbre del clima	32	1290	68,73325	0,06584	65,426
20	emisión de gases	60	22981	128,87483	1,17291	59,77
21	impacto del cambio	37	8238	79,47282	0,42045	56,653
22	temperatura media	67	33462	143,91023	1,70785	53,515
23	impacto del cambio climático	37	10374	79,47282	0,52947	52,615
24	emisión neta	25	914	53,69785	0,04665	52,26
25	fenómeno meteorológico	41	14372	88,06447	0,73352	51,378
26	fenómeno extremo	24	1256	51,54993	0,0641	49,384
27	reducción de emisiones	43	19035	92,3603	0,97152	47,355
28	aumento de las temperaturas	26	4026	55,84576	0,20548	47,156
29	emisión de gases	68	41596	146,05815	2,12299	47,089
30	transición ecológica	21	477	45,10619	0,02435	45,01
31	emisión global	23	2589	49,40202	0,13214	44,519
32	objetivo climático	21	728	45,10619	0,03716	44,454
33	consecuencia del cambio climático	25	4746	53,69785	0,24223	44,032
34	vuelo corto	21	1298	45,10619	0,06625	43,242
35	plan climático	19	143	40,81036	0,0073	41,507
36	inseguridad alimentaria	27	8356	57,99368	0,42648	41,356
37	incendio forestal	75	58096	161,09355	2,96513	40,88
38	política climática	20	1669	42,95828	0,08518	40,508
39	consecuencia del cambio	25	6865	53,69785	0,35038	40,506
40	aumento de la temperatura global	19	883	40,81036	0,04507	40,007
41	militancia política	26	8309	55,84576	0,42408	39,918
42	activista climático	18	92	38,66245	0,0047	39,477
43	presupuesto de carbono	18	299	38,66245	0,01526	39,066
44	aumento del nivel	26	9017	55,84576	0,46021	38,93
45	baja emisión	25	8017	53,69785	0,40917	38,816
46	energía limpia	45	30860	96,65613	1,57505	37,924
47	adaptación al cambio climático	25	8787	53,69785	0,44847	37,762
48	extensión del hielo	17	541	36,51454	0,02761	36,507
49	capa de hielo	26	10965	55,84576	0,55964	36,448
50	sur global	17	853	36,51454	0,04354	35,949
51	dióxido de carbono	65	58974	139,61441	3,00994	35,066
52	temperatura media mundial	16	482	34,36662	0,0246	34,517
53	negociación climática	16	652	34,36662	0,03328	34,228
54	subida del nivel	18	3391	38,66245	0,17307	33,811
55	aumento de temperaturas	16	1086	34,36662	0,05543	33,509
56	aumento del nivel del mar	18	3621	38,66245	0,18481	33,476
57	impacto climático	16	1177	34,36662	0,06007	33,362
58	zona de bajas emisiones	15	197	32,21871	0,01005	32,888

59	subida del nivel del mar	16	2026	34,36662	0,1034	32,052
60	energía renovable	127	148618	272,78506	7,58523	31,89
61	tolerancia humana	14	85	30,0708	0,00434	30,937
62	producción de combustibles fósiles	14	155	30,0708	0,00791	30,827
63	extensión del hielo marino	14	268	30,0708	0,01368	30,652
64	justicia climática	15	1663	32,21871	0,08488	30,62
65	ruta aérea	17	5006	36,51454	0,2555	29,88
66	calentamiento del planeta	15	2344	32,21871	0,11963	29,669
67	mosquito tigre	15	2665	32,21871	0,13602	29,241
68	producción de combustibles	14	1292	30,0708	0,06594	29,149
69	sistema alimentario	17	5756	36,51454	0,29378	28,996
70	emisión mundial	14	1482	30,0708	0,07564	28,886
71	emisión de dióxido de carbono	19	8856	40,81036	0,452	28,795
72	calidad del aire	37	35205	79,47282	1,79681	28,773
73	plataforma de hielo	14	1766	30,0708	0,09013	28,502
74	vuelo doméstico	15	3275	32,21871	0,16715	28,461
75	industria de los combustibles fósiles	13	352	27,92288	0,01797	28,412
76	economía circular	18	7881	38,66245	0,40223	28,285
77	masa de aire frío	14	1993	30,0708	0,10172	28,202
78	pérdida de biodiversidad	17	6610	36,51454	0,33736	28,051
79	alta temperatura	67	82670	143,91023	4,21935	27,764
80	transición justa	13	873	27,92288	0,04456	27,689
81	sequía extrema	13	1322	27,92288	0,06747	27,095
82	aumento global	13	1373	27,92288	0,07008	27,029
83	aire frío	23	17110	49,40202	0,87327	26,906
84	refugio climático	12	53	25,77497	0,00271	26,703
85	récord de calor	12	299	25,77497	0,01526	26,373
86	riesgo meteorológico	12	425	25,77497	0,02169	26,207
87	nivel del mar	54	68651	115,98735	3,50384	25,975
88	temperatura central	12	728	25,77497	0,03716	25,816
89	efecto del calentamiento	13	2438	27,92288	0,12443	25,722
90	lluvia torrencial	20	14142	42,95828	0,72179	25,531
91	temperatura extrema	18	12066	38,66245	0,61583	24,546
92	lucha climática	11	99	23,62705	0,00505	24,503
93	temperatura récord	11	501	23,62705	0,02557	24,013
94	era preindustrial	11	657	23,62705	0,03353	23,828
95	contaminación atmosférica	20	16670	42,95828	0,85081	23,751
96	aire cálido	12	2632	25,77497	0,13433	23,604
97	estrés térmico	11	1919	23,62705	0,09794	22,43
98	alternativa ferroviaria	10	64	21,47914	0,00327	22,406
99	grado centígrado	27	32267	57,99368	1,64686	22,288
100	grado celsius	16	11540	34,36662	0,58898	22,257

Fuente: Elaboración propia a partir de la herramienta de extracción de *multi-word terms* de Sketch Engine (2024).

A diferencia de las palabras clave, observamos en esta lista que en las expresiones multipalabra no aparecen nombres propios, puesto que se trata de palabras con contenido léxico. Si nos quedamos solo con los elementos léxicos que componen las combinaciones léxicas, tendríamos un total de 215 elementos (solo de la lista reducida de 100 ítems que hemos incluido en este trabajo), contabilizados de la siguiente manera (% de aparición en las combinaciones léxicas):

- **8,72%:** *climático/a*
- **5,05%:** *temperatura/s*
- **4,13%:** *cambio, emisión/es*
- **3,21%:** *aumento*
- **2,75%:** *global*
- **2,29%:** *hielo, nivel*
- **1,83%:** *aire, combustible/s, efecto, extremo/a*
- **1,38%:** *calentamiento, calor, carbono, fósil/es, gas/es, impacto, transición.*
- **0,92%:** *alimentario/a, baja/s, consecuencia, dióxido, energía, extensión, fenómeno, frío, grado, invernadero, lucha, mar, marino, media, mundial, política, preindustrial, producción, récord, subida, vuelo.*
- **0,46%:** *acción, activista, adaptación, aérea, alta, alternativa, atmosférica, biodiversidad, calidad, cálido, capa, celsius, centígrado, central, circular, clima, contaminación, corto, crisis, cumbre, doméstico, ecológica, economía, emergencia, energética, era, estrés, fenómeno, ferroviaria, forestal, herramienta, humana, incendio, inseguridad, justa, justicia, limpia, lluvia, masa, meteorológico, militancia, mosquito, negociación, neta, nivel, objetivo, ola, pérdida, plan, planeta, plataforma, presupuesto, reducción, refugio, renovable, riesgo, ruta, sequía, sistema, sur, térmico, tigre, tolerancia, torrencial, zona.*

Hemos comparado los elementos léxicos de las dos listas (palabras clave y expresiones multipalabra) y los que aparecen exclusivamente en la lista de palabras clave son las siguientes: *borrasca, carbono, CO2, countdown, covid, deforestación, derretimiento, descarbonización, deshielo, ecodidio, ecologismo, eucalipto, geotermia, gigatonelada, glaciario, lúpulo, metano, meteorología, meteorólogo, mitigación, negacionismo, OC, pandemia, percebe, posidonia, precipitación, quebrantahuesos, récord, resiliencia, sequía, veintisiete y ZBE.*

En el caso de las expresiones multipalabras, añadimos a continuación los elementos léxicos exclusivos de esta lista: *acción, activista, adaptación, aérea, aire, alimentario/a, alta, alternativa, atmosférica, aumento, baja/s, biodiversidad, calidad, cálido, calor, cambio, capa, carbono, celsius, centígrado, central, circular, climático/a, consecuencia, contaminación, corto, crisis, cumbre, doméstico, ecológica, economía, efecto, emergencia, energética, energía, era, estrés, extensión, extremo/a, fenómeno, ferroviaria, forestal, frío, gas/es, global, grado, herramienta, humana, impacto, incendio, industria, inseguridad, justa, justicia, limpia, lluvia, lucha, mar, marino, masa, media, meteorológico, militancia, mosquito, mundial, negociación, neta, nivel, nivel, objetivo, pérdida, plan, planeta, plataforma, política, preindustrial, presupuesto, producción, récord, reducción, refugio, renovable, riesgo, ruta, sistema, subida, sur, térmico, tigre, tolerancia, torrencial, transición, vuelo y zona.*

Para terminar la comparación de las dos listas, indicamos los elementos léxicos que aparecen en ambas listas: *calentamiento, calor, combustible, dióxido, emisión, fósil, hielo, invernadero, ola, supercontinente y temperatura.*

En la lista de palabras clave aparece el ítem *climate* y *countdown*, lo hemos mantenido en el corpus bajo su forma original, en inglés. En el caso de *climate*, en la lista de combinaciones léxicas aparece en español, *clima*. En cuanto a las siglas ZBE y OC recogidas en la lista de palabras clave, también las encontramos en las expresiones multipalabra: *zona de bajas emisiones*

y ola de calor.

4. Conclusiones

Tal y como nos propusimos al comienzo de este trabajo, hemos realizado una revisión de las palabras clave de un corpus textual sobre cambio climático que nosotros hemos confeccionado. Si bien en este estudio hemos indagado principalmente en las palabras clave y las expresiones multipalabra, este corpus nos servirá para futuras investigaciones, pues resultaría interesante cruzar los datos obtenidos en las listas de palabras por frecuencia con otros corpus de distinta naturaleza, como, por ejemplo, con datos obtenidos a partir de un trabajo de campo propio de los estudios sociolingüísticos. Asimismo, se podría comparar los datos con otros corpus textuales procedentes de otras fuentes de distinto género. Si bien el corpus confeccionado no se trata de una muestra representativa de las noticias publicadas en España, pues solo hemos recopilado noticias durante un mes y de determinadas marcas, no se trata de un corpus cerrado, podemos seguir aumentando tanto el rango de fechas como las marcas y el tipo de género de los textos. No obstante, el número total de palabras que componen el corpus de este trabajo es de 405.463 aproximadamente.

En cuanto al programa de confección y gestión de corpus usado en este trabajo, Sketch Engine, facilita la gestión de tu propio corpus textual, además de ofrecer una multitud de posibilidades de estudios de diferente naturaleza y temática, esto es, adaptado a los intereses del investigador.

En lo que respecta a los resultados obtenidos, nos ha servido para observar e indagar desde los textos escritos sobre cambio climático publicados en diferentes marcas de medios de comunicación cómo las distintas esferas de la sociedad (economía, política y educación, entre otras) están relacionadas con esta causa. En futuros estudios sería interesante realizar un análisis comparativo de las palabras más frecuentes en los diferentes medios donde se han publicado las noticias que componen el corpus textual del presente trabajo, así como también explorar el contenido de este corpus a través de los contextos de aparición.

Concluyendo, el presente trabajo nos ha servido como lingüista para profundizar en esta temática desde una dimensión lingüística y describir, de manera objetiva, la sociedad a través de estos textos.

5. Referencias

- Barkemeyer, R., Figge, F., Hoepner, A., Holt, D., Kraak, J. M. y Yu, P. S. (2017). Media coverage of climate change: An international comparison. *Environment and Planning C: Politics and Space*, 35(6), 1029-1054. <https://doi.org/10.1177/0263774X16680818>
- Boykoff, Maxwell T. y J. Timmons Roberts. 2007. Media coverage of climate change: Current trends, strengths, weaknesses. *Human Development Report 2007/2008. Fighting climate change: Human solidarity in a divided world*, United Nations Development Programme. <https://hdr.undp.org/content/media-coverage-climate-change>
- Gillings, M. y Dayrell, C. (2023). Climate change in the UK press: Examining discourse fluctuation over time. *Applied Linguistics*, 45(1), 111-133. <https://doi.org/10.1093/applin/amad007>
- Grundmann, R. y Ramesh, K. (2010). The Discourse of Climate Change: A Corpus-based

- Approach. *Critical Approaches to Discourse Analysis across Disciplines*, 4(2), 113-133. <https://shre.ink/DTOC>
- Kilgarriff, A. (2009). Simple maths for keywords. En M. Mahlberg, V. González-Díaz y C. Smith (Eds.), *Proceedings of Corpus Linguistics Conference. CL2009*. University of Liverpool.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý y V. Suchomel. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7-36. <https://shre.ink/DTOC>
- Liu, M. y Huang, J. (2022). “Climate change” vs. “global warming”: A corpus-assisted discourse analysis of two popular terms in The New York Times. *Journal of World Languages*, 8(1), 34-55. <https://doi.org/10.1515/jwl-2022-0004>
- Sánchez-Saus Laserna, M. y Álvarez Torres, V. (2022). ¿De qué hablamos cuando divulgamos sobre lingüística? Análisis de un corpus de textos divulgativos y aplicaciones al estudio terminológico de la semántica léxica. *ELUA (Estudios de Lingüística. Universidad de Alicante)*, 38, 73-98. <https://doi.org/10.14198/ELUA.22384>
- Statista Search Department. (2023). *Ranking de las principales marcas de medios de comunicación online según el porcentaje de población que las usaba de forma semanal en España en 2022* [Infographic]. <https://shre.ink/DTOu>
- UNICEF Comité Español. (2017). *Glosario sobre cambio climático para el aula. Nuestro planeta, nuestros derechos: educación, derechos de infancia y cambio climático*. <https://www.unicef.es/educa/biblioteca/glosario-cambio-climatico>
- United Nations Framework on Climate Change. (s.f.). *¿Qué es el Protocolo de Kyoto?* https://unfccc.int/es/kyoto_protocol
- United Nations Framework on Climate Change. (s.f.). *Conferencia de las Partes (COP)*. <https://shre.ink/DTOj>
- Volkanovska, E., Tan, S., Duan, C., Changxu, D., Bartsch, S. y Stille, W. (2023). The InsightsNet Climate Change Corpus (ICCC). *Datenbank Spektrum*, 23, 177-188. <https://doi.org/10.1007/s13222-023-00454-1>
- Willis, R. (2017). Taming the Climate? Corpus analysis of politicians’ speech on climate change. *Environmental Politics*, 26(2), 212-231. <https://doi.org/10.1080/09644016.2016.1274504>

AUTORA:

Vanesa Álvarez Torres
Universidad de Cádiz.

Vanesa Álvarez Torres es Licenciada en Lingüística, Maestra en la especialidad de Lengua Extranjera y Doctora en Lingüística por la Universidad de Cádiz, con Premio Extraordinario de Doctorado. Ha sido personal técnico de apoyo del Programa Estatal de Promoción al Talento y su Empleabilidad en I+D+i del MINECO en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz (ILA). En la actualidad, es Profesora Ayudante Doctora del área de Lingüística General (departamento de Filología) de la Universidad de Cádiz

y miembro del grupo de investigación “Semaínein” y del Instituto Universitario de Investigación en Lingüística Aplicada.

vanesa.alvarez@uca.es

Orcid ID: <https://orcid.org/0000-0003-0477-4065>

Scopus ID: <https://www.scopus.com/authid/detail.uri?authorId=57999568400>

Google Scholar: <https://scholar.google.com/citations?user=LFmGV0QAAAAJ>

ResearchGate: <https://www.researchgate.net/profile/Vanesa-Alvarez-Torres>