Mapping the Discourse of Climate Change in four Western Countries

Supplementary information

Parsing and filtering

Ideally, a corpus building on news resources should have certain properties which we tried to achieve:

- ensure a standard format
- identify article features such as headlines, text boxes, sub-headings, captions of graphics, hypertext links
- filter out duplicate articles and duplicated sections within articles
- filter out articles where the main topic searched for is not evident
- store individual articles so that they can be retrieved easily
- build sub-corpora by grouping articles according to criteria such as date, publication-type, or within a set of specific publications

Accordingly, the process involved parsing the downloads to arrive at individual articles, identifying headlines and meta-data such as publication date, byline, or section of the paper. Articles were dated and sorted into one folder for each year for each of the newspapers for each country. Metadata in fields such as LENGTH, HEADLINE, BYLINE etc in the header were used for filtering texts as described below. French and German articles in most cases use field-names in those languages.

Besides this parsing process, extraneous mark-up needed to be removed. Each data source gets passed to the data provider such as LexisNexis in a slightly different format, and stored with such inconsistencies retained. There are web-links embedded in the text, HTML references to other authors or to captions, insertions such as "block-time updated-timeUpdated at 7.41am GMT" which look like vestiges of internal edit mark-up. Each newspaper will mark up texts with codes as in <SECTION: US NEWS> and <JOURNAL-CODE: WEBGNS> to suit their publishing categories.

We then filtered the corpus as described below. Texts were removed if they met any of these criteria:

- below 400 words in length
- undated
- from a regional section of a newspaper or where a UK newspaper had a section explicitly called Australia News or US News, flagged in the header as explained above.
- duplicates or near-duplicates of others
- not containing enough climate-change relevance

Duplicates were removed using the procedure outlined at https://lexically.net/downloads/version9/HTML/duplicates_head.html, which looks for the same file-name or with a very minor variation, such as ST000083.txt and ST000083 (3).txt or ST000083 – Copy.txt

Where the file-names are different, the procedure combined a) detecting 100-character matches with b) comparison of the lexis of the texts in question. For b), the settings required a maximum length difference of 8% between texts being compared, a maximum difference in the number of types of 5%, and comparison was only within the same folders (based on year of publication).

Relevance: it was obvious at an early stage of reading sample texts in the processes outlined above that many texts were not really about climate change. For example, at the end of a year a journalist writes a general report on the year just finished and makes mention of weather events in certain places and months. Other examples are expressions like ‘a climate of tension’, ‘ a climate of uncertainty’, etc. where the term climate is typically metaphorical. We felt it important to filter our corpora drastically to exclude such cases, trying to reach a point where we could feel confident the texts really concerned climate change, including its social, economic, or political dimensions.

Accordingly, we used the relevance filter in WordSmith (https://lexically.net/downloads/version9/HTML/relevance_check.html) which searches each text for relevant key words and allocates a score based on the number of key words found in the text, moderated by the number of segments of the text each key word is present in. For the current purpose the segments were fifths of the text, and the key words were terms obtained in our previous work on climate change.

The resulting scores do not usually show a clear cut-off point. A text about politics in Philadelphia mentions climate change enough to get a fairly high score but is it really a text about Republicans disagreeing with Democrats in general with some mention of climate change? Similarly a text on US states’ water resources scores high, but how climate change-ish is the whole water resource article? Accordingly, after running the relevance check we sampled texts at 15 bands within the range of scores. Those in the lowest 5 bands were very easily eliminated. After much reading by both of the present authors (it would have been preferable to have engaged more readers) we were able to determine a cut-off point where most of the texts were clearly on the topic of climate change. We thus kept the top 5 bands only. Table 1 shows the number of articles after initial download and after cleaning and filtering.

	Initial download	After parsing
Le Monde/ Monde Diplomatique	7630	2818
Le Figaro/Figaro Economique	5379	598
Les Echos	3439	775
Le Temps	3267	1362
La Tribune	2908	1393
Sud Oest	2095	259
L’Humanité	1044	307
FR Total	**25762**	**7512**
Frankfurter Allgemeiner Zeitung	9217	2319
Die Welt/ Welt am Sonntag/ Welt Online	6514	2807
TAZ, Die Tageszeitung	5553	2131
Stuttgarter Zeitung	4931	917
Der Spiegel/Spiegel Online	8505	2309
Die Zeit	3636	1277
Berliner Zeitung	3618	933
DE Total	**48203**	**12693**
The Guardian/ Observer	29224	8826
Financial Times	16320	2717
Daily Mail/Mail on Sunday	15035	1901
Independent/Independent on Sunday	13660	4516
Daily Telegraph/Sunday Telegraph	12202	2587

The Times/Sunday Times	11260	2196
UK Total	**97701**	**22743**
New York Times	16820	8045
Washington Post	8875	7555
St Louis Post Dispatch	2099	452
Star Tribune	1879	471
Philadelphia Inquirer	1695	538
USA Today	1554	622
US Total	**36282**	**17683**

Table 1: Corpus of news articles 2005-2022 initial download and parsed corpus

Basic Statistics

After cleaning and filtering, we found these basic statistics:

France: The 8.6 million-word corpus constructed with these cleaning up and filtering operations done has 113 thousand word types, with an average sentence length of 16.4 words. The most frequent words are grammatical; the first ten lexical words are PAYS, CLIMATIQUE, CLIMAT, EMISSIONS, FRANCE, CARBONE, RECHAUFFEMENT, EFFET, MONDE, ETATS.

Germany: The 14.6 million-word corpus constructed with these cleaning up and filtering operations done has 332 thousand word types, with an average sentence length of 12.9 words. The most frequent words are grammatical; the first ten lexical words are PROZENT, JAHREN, GRUENEN, MENSCHEN, DEUTSCHLAND, KLIMAWANDEL, KLIMATSCHUTZ, CHINA, JAHRE, USA.

UK: The 33.7 million-word corpus constructed with these cleaning up and filtering operations done has 182 thousand word types, with an average sentence length of 16.1 words. The most frequent words are grammatical; the first ten lexical words are CLIMATE, SAID, CHANGE, ENERGY, PEOPLE, GLOBAL, WORLD, CARBON, GOVERNMENT, UK.

US: The 25.9 million-word corpus constructed with these cleaning up and filtering operations done has 131 thousand word types, with an average sentence length of 18.0 words. The most frequent words are grammatical; the first ten lexical words are SAID, CLIMATE, TRUMP, NEW, CHANGE, PRESIDENT, ENERGY, YEAR, PEOPLE, STATES.

Yearly coverage

Yearly news coverage per newspaper shows a huge variation, across newspapers, and within newspapers over time (see Table 2:). Elite newspapers publish many more articles, and especially so after 2016. In the US, the growth is driven by the *New York Times* and *Washington Post.* The *Washington Post* shows the fastest growth, from 315 articles in 2017 to 1417 articles in 2019 (when the average number of articles in our US sample was 406 for that year).

In the UK it is the *Guardian* which has the highest coverage. In 2015 *The Guardian* published 1035 articles (average 157); in France *Le Monde* published 400 articles compared to an average of 122. In Germany, *FAZ*, *Die Welt*, *Spiegel* and *taz* drive the increasing attention in 2007, 2019 and 2021, but the difference to other papers is not so pronounced.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Sum
Berliner_Zeitung	7	16	70	21	40	33	24	13	21	32	52	51	61	50	135	70	133	104	
Frankfurter Allgemeine	44	50	203	121	122	89	73	45	67	67	119	61	124	167	140	142	371	314	
SPIEGEL	34	68	266	133	178	110	76	67	80	67	84	45	92	145	375	222	267	136	
Stuttgarter Zeitung	12	20	82	49	51	38	22	19	30	24	45	35	53	30	108	54	119	126	
TAZ	31	44	104	70	76	55	64	36	50	51	92	84	126	93	307	233	329	286	
WELT	49	81	262	129	124	91	74	103	73	83	154	128	216	237	322	215	295	171	
Zeit				18	87	53	54	32	22	31	54	36	63	92	246	146	208	135	
ECHOS	34	38	50	29	45	19	17	12	25	25	76	22	19	30	38	98	136	62	
FIGARO	28	40	43	25	33	18	9	16	11	25	94	16	49	27	48	28	54	34	
LA TRIBUNE	10	27	18	17	19	9	2	28	23	26	174	99	85	114	90	102	235	315	
LE TEMPS	43	55	122	83	71	62	51	38	51	42	39			77	150	101	156	221	
L'HUMANITÉ	6	8	11	10	8	6	4	3	10	16	39	10	18	32	26	23	38	39	
MONDE	51	36	91	67	150	120	112	100	85	113	400	93	144	191	203	190	288	384	
SUD OUEST	9	6	6	2	10	11	7	5	10	5	31	10	18	18	29	33	30	19	
FINANCIAL TIMES	96	104	154	131	183	93	86	80	72	78	107	75	85	119	252	233	419	350	
GUARDIAN and OBSERVER	203	238	366	310	478	307	300	381	429	440	1035	588	433	467	793	586	811	661	
INDEPENDENT and INDEP on SUNDAY	146	198	195	112	185	119	92	84	113	97	140	199	232	178	273	306	930	917	
MAIL and SUNDAY MAIL	19	46	83	44	41	29	32	93	181	250	274	253	224	144	37	26	55	70	
TELEGRAPH and SUNDAY TELEGRAPH	37	91	148	87	116	185	165	202	210	154	160	155	135	105	121	117	225	174	
TIMES and SUNDAY TIMES	77	152	210	140	308	160	105	84	93	70	73	56	65	72	138	50	221	122	
NEW YORK TIMES	102	135	282	248	319	190	146	160	216	331	390	480	718	720	827	806	1091	884	
PHILADELPHIA INQUIRER	20	15	28	34	22	16	15	10	15	13	18	28	32	32	72	62	58	48	
ST. LOUIS POST-DISPATCH	6	11	24	35	28	15	8	9	15	25	33	41	64	48	41	21	14	14	
STAR TRIBUNE	2	14	30	32	17	18	13	14	28	19	24	18	30	27	45	36	55	49	
USA TODAY	8	20	49	40	38	21	12	14	22	11	8	12	21	9	33	26	67	211	
WASHINGTON POST	61	78	219	201	256	146	87	107	112	140	174	205	315	1160	1417	1074	1255	548	

Table 2: Yearly news coverage, number of articles

Named Entity Recognition (NER)

A major purpose of our research was to identify names of people and organisations central to debates and activities around climate change. To do that, a Python program was written which uses resources from spacy (https://spacy.io/). It can locate named entities in plain text (en_core_web_sm for English, de_core_news_sm for Gerrman and fr_core_news_sm for French) and list the results in a text file for each text. It gives useful but rather rough and ready results. It classifies words according to such categories like Person, Organization, Place or Geographical Political Entity (GPE). We are only interested in Persons and Organizations.

The application of NER software needed close scrutiny.

Let us consider an example of output for a 29 December 2022 article published in *The Times*. Under the headline

Cameron's climate climbdown cost us dear; A decade on, the former PM's U-turn on renewables looks increasingly misguided and out of kilter with public opinion

the article starts as follows:

T is the season of love and peace, and goodwill to all mankind. Except David Cameron. Our unmissed former prime minister is part of the reason Britain faces a bleak midwinter over energy bills. You'll pay more to heat and light your home in 2023 because in 2013 Cameron demanded an end to the "green crap" added to energy bills to support energy efficiency and renewable electricity generation; he later banned new onshore wind turbines. As a result, we're more reliant than we should have been on gas prices that are partly determined by Vladimir Putin. Thanks, Dave.

I'm not rehearsing this tale solely out of festive spite towards our former PM. There are useful lessons in contemplating just how wrong Cameron was.

The context for Cameron's 2013 shift on green energy was a squeeze on the cost of living. Table 3: shows named entities in order of appearance.

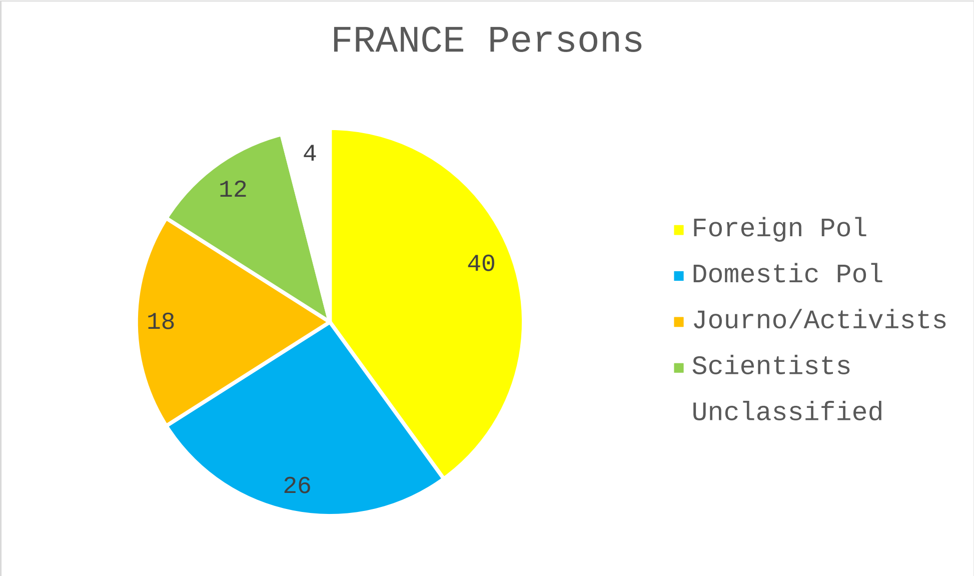
Cameron	PERSON
Us	GPE
A decade	DATE
David Cameron	PERSON
Britain	GPE
2023	DATE
2013	DATE
Cameron	PERSON
Vladimir Putin	PERSON
Dave	PERSON
Cameron	PERSON
Cameron	ORG
2013	DATE

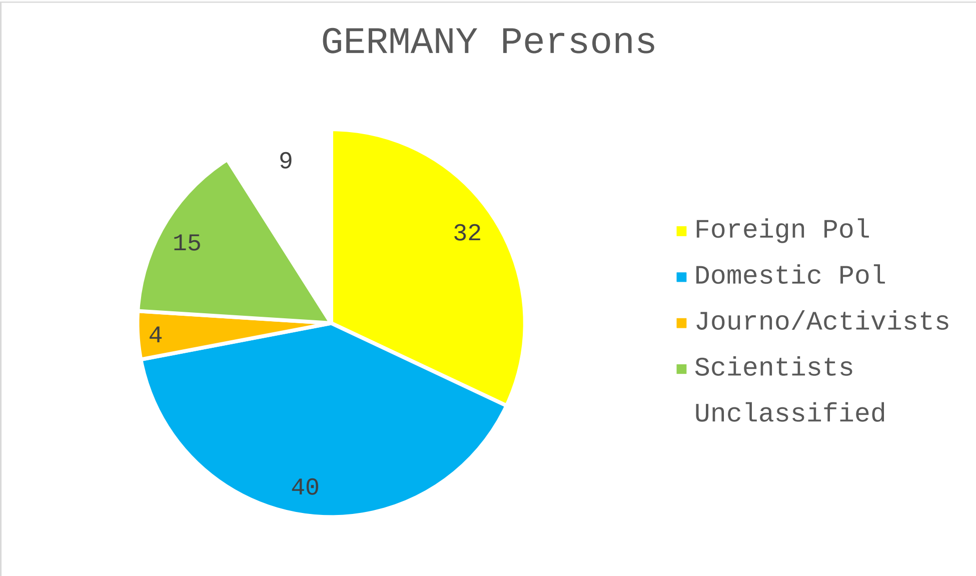
Table 3: Named entities classification, example

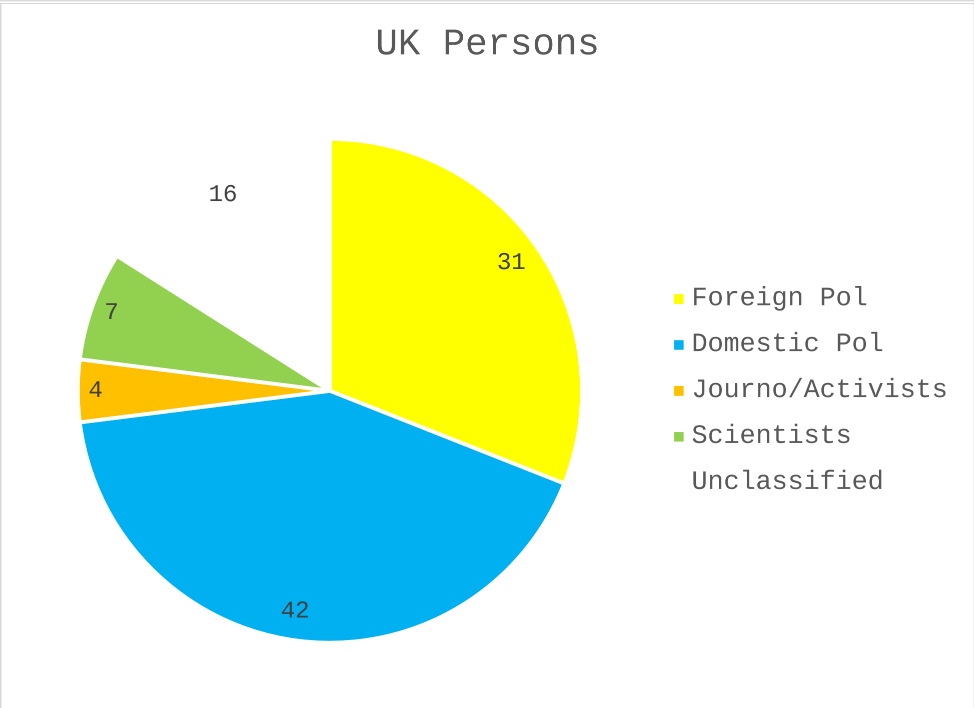
The process mistook the pronoun 'us' for 'US' in row 2 (GPE). *How wrong Cameron was* has Cameron correctly identified as a person but *Cameron's 2013 shift on green energy* mis-identifies him as an organization. The French and German data use slightly different forms but perform to a similar level. This means we had to check manually for misallocations and correct the dataset accordingly. While we did most of our analysis before the advent of ChatGPT we are not confident that this would have made things easier.

Top 100 named entities

We also performed an analysis of the top 100 persons and organisations. The pie charts below (Figure 1a-d) show the proportion of the types of top 100 persons for each country.


FRANCE Persons
4
12
40
18
26
Foreign Pol
Domestic Pol
Journo/Activists
Scientists
Unclassified


GERMANY Persons
9
15
32
4
40
Foreign Pol
Domestic Pol
Journo/Activists
Scientists
Unclassified


UK Persons
16
31
7
4
42
Foreign Pol
Domestic Pol
Journo/Activists
Scientists
Unclassified

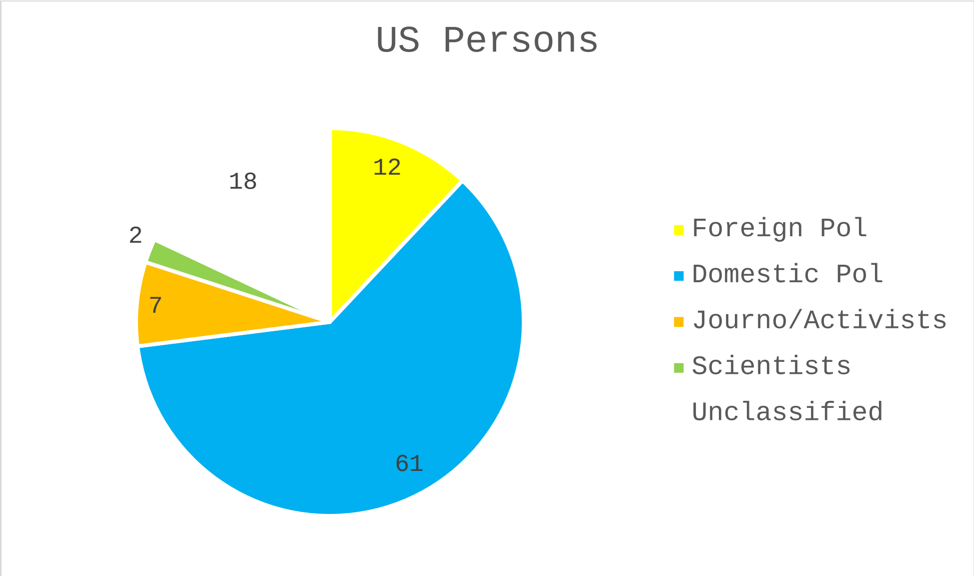
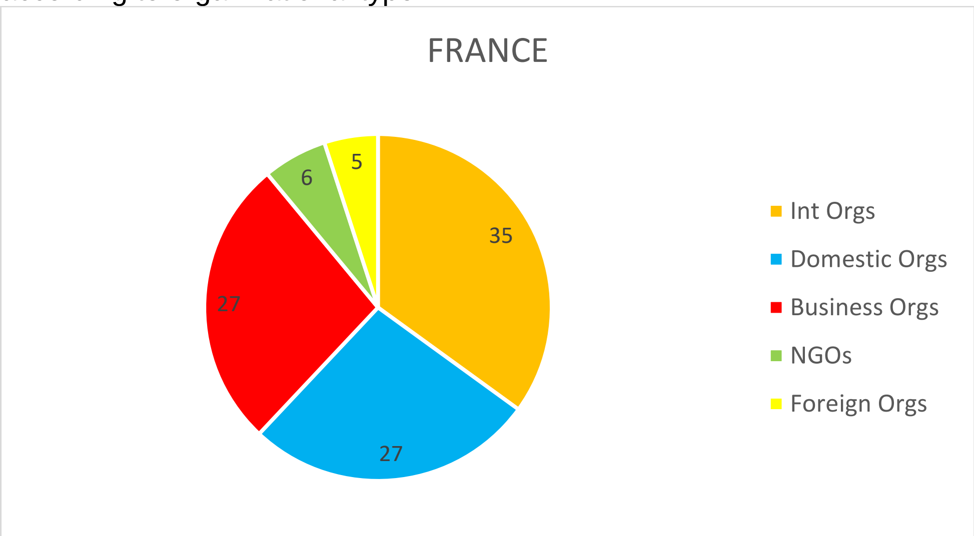
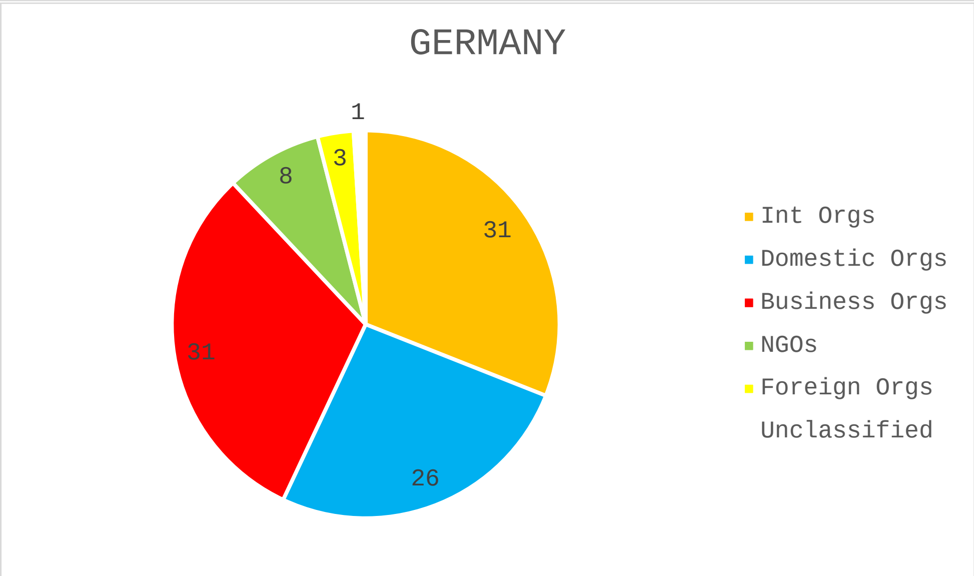
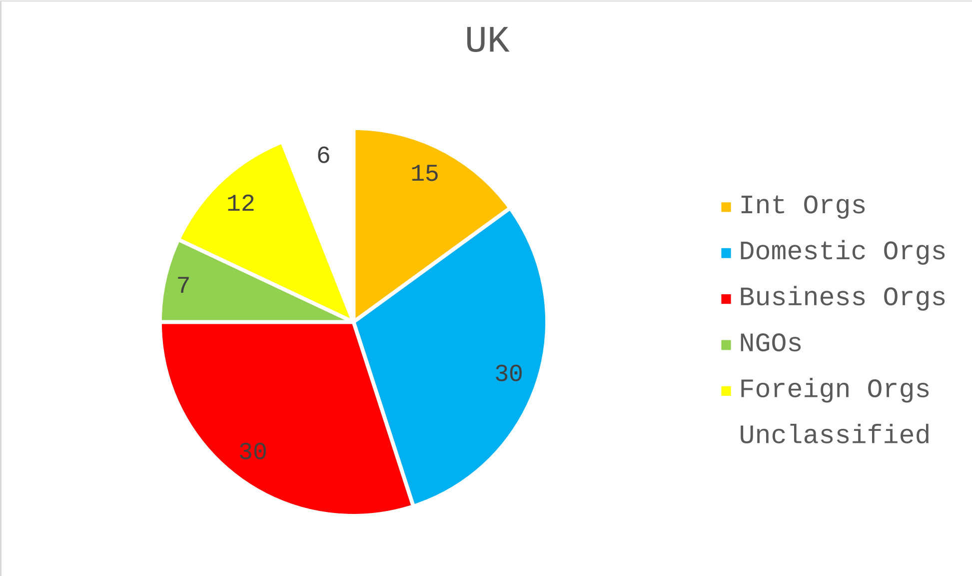


Figure 1. a-d: Pie charts of the proportions of different actor categories, based on the top 100 mentions

The pie charts below (Figure 2a-d) detail the proportion of the top 100 organizations mentioned in the press according to organizational type.







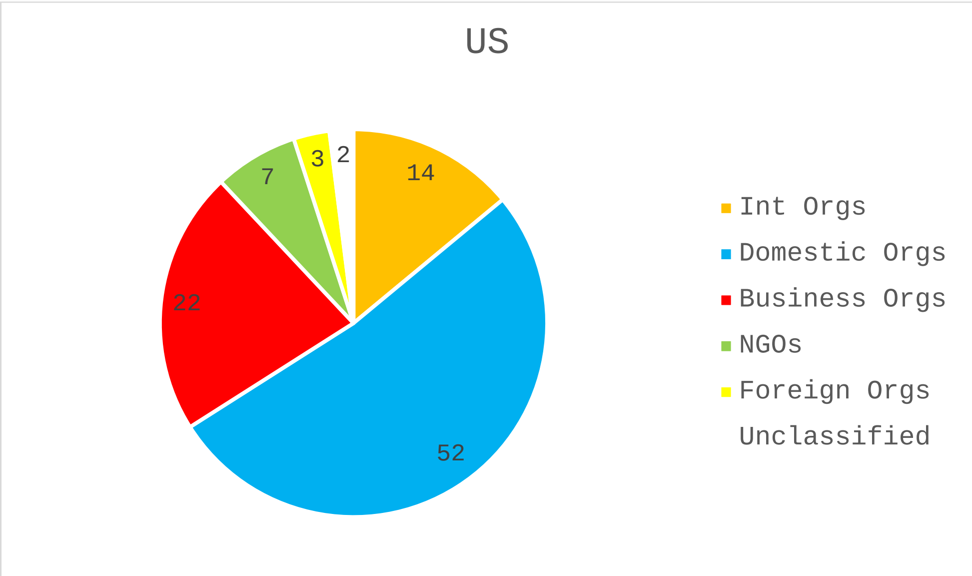


Figure 2. a-d: Pie charts of the proportions of different organization types, based on the top 100 mentions

Collocation analysis

After the Named Entity Recognition process we performed a collocation analysis, using key actors identified using NER. A concordance of the key actor term gives a standard context span of 5 words to left and right, these words are selected for study. A term such as *Merkel* will be found with true collocates such as *Chancellor* but also with unspecific words like *of, was* and *the.* Our aim is to find identify terms like *Chancellor* that are strongly linked -- which implies rarely found apart from *Merkel*. We then created multiple sets of ten words and studied them statistically, in this case using a Z score, using the joint frequencies and frequencies in other contexts. We identified the top words as those which are most strongly linked.

Analysis of peaks

The tables below show the time series for the top 25 persons and organisations in Germany, France, the UK, and the US. Yearly maxima are highlighted.

Colour coding Persons:

Domestic pol. | Foreign pol. | Scientist | Activist/journalist

Colour coding Organisations:

International org. | Domestic org. | Foreign org. | Business org. | Env. NGO

Persons	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Angela Merkel	230	241	1037	403	664	356	237	287	329	301	448	260	838	250	343	268	285	51
Barack Obama			23	635	1112	379	225	753	488	362	462	309	201	92	31	57	32	10
Donald Trump											40	1096	1794	793	286	470	168	61
Sigmar Gabriel	158	780	613	406	253	31	68	41	68	962	401	291	221	29	36	8	4	5
George W. Bush	1328	560	588	383	142	74	29	50	15	15	32	19	38	14	1	7	5	2
Robert Habeck						7	9	6	8	40	6	169	247	280	207	182	368	349
Al Gore	5	605	330	109	74	101	20	14	30	30	23	19	36	24	6	8	2	8
Winfried Kretschmann	10	14	3	12	4		115	22	214	83	56	229	126	32	68	100	126	55
Cem Özdemir			2	151	36	34	9	17	48	78	23	134	344	23	36	19	52	80
Greta Thunberg														6	633	276	74	47
Jürgen Trittin	148	105	43	70	57	38	34	77	151	83	13	16	74	20	7	12	11	6
Annalena Baerbock							2				6	10	35	151	103	78	437	128
Katrin Göring-Eckardt		3	7		12	4		44	156	156	7	74	279	68	40	9	11	2
Olaf Scholz			2	13	4	4	36	8		15	20	12	9	31	51	68	312	271
Christian Lindner	5					4		22	5	15	4		257	72	124	64	179	101
Wladimir Putin	56	14	74	18	16	9	2	11	8	103	50	56	97	58	34	22	57	162
Barbara Hendricks			1						3	279	212	161	106	20	2		2	1
Stefan Rahmstorf	235	38	112	13	19	31	106	33	33	5	24	2	20	8	11	15	4	4
Renate Künast	10	105	44	58	66	187	63	22	28	30	4		7	5	4	3	19	1
Emmanuel Macron													221	116	131	79	42	62
Armin Laschet					5	20		14	28		19		35	19	18	60	398	5
Tony Blair	296	143	107	15	18	7	5	3			1		3		1	1		1
Claudia Roth	5	164	56	38	36	63	2	61	60	43	3	8	12	3	9	4	9	5
Norbert Röttgen			1	2	112	172	106	113	8	8	1	2	2	1	2	11	17	
Peter Altmaier			1					146	153	23	4	4	21	62	40	66	32	4

Table 4: NER results DE Persons, mentions per million words, yearly maximum values highlighted.

Organisations	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Die Grünen	878	682	435	781	603	970	812	535	1320	866	273	791	1592	1066	1239	1036	1584	658
EU	363	455	568	589	599	399	471	480	478	766	429	470	462	531	453	741	459	414
SPD	322	252	317	388	345	386	435	232	498	530	236	385	477	566	678	381	775	257
CDU	230	84	212	561	282	456	325	251	402	216	115	369	424	322	606	349	685	201
FDP	82	143	58	91	288	178	110	127	302	136	23	37	778	130	220	151	581	223
IPCC	82	70	286	106	109	349	142	245	387	347	62	45	66	107	73	58	80	80
UN	138	164	212	146	203	165	167	143	128	143	224	150	131	114	112	105	79	108
Bundestag	87	49	70	86	73	101	50	33	116	156	50	76	141	98	156	163	211	107
EU-Kommission	61	126	143	88	60	49	34	105	85	246	73	74	66	81	91	167	81	81
AfD									25	100	9	460	179	177	412	127	129	33
CSU	15	101	60	91	66	32	63	19	91	40	53	130	266	261	143	62	121	30
RWE	87	84	69	100	52	97	90	124	38	98	125	105	100	104	62	36	57	90
Greenpeace	153	49	89	58	60	122	108	55	73	88	85	56	50	34	49	37	52	59
PIK	56	101	61	43	77	74	45	88	88	20	98	54	28	70	44	41	48	43
Weltbank	72	105	51	65	66	40	50	152	106	53	40	52	45	27	38	34	21	32
EU-Staaten	41	63	51	50	42	13	25	55	58	100	66	39	47	53	40	62	35	34
WWF	66	112	69	48	57	59	59	99	23	38	47	41	17	34	19	15	23	22
UN-Klimakonferenz	20	42	25	33	88	45	54	44	35	33	82	16	47	58	30	11	31	43
Vattenfall	20	91	70	85	34	38	41	8	91	78	46	21	20	28	9	9	4	4
Europaparlament	10	14	26	35	27	36	14	8	68	118	14	29	20	37	88	87	29	35
Nato	107	10	18	45	27	7	18	11	8	10	10	52	54	29	9	57	95	51
WTO	36	14	64	32	29	7	50	17	10	13	46	39	59	85	22	19	32	24
EPA		38	21	12	14	20	9	8	8	10	56	111	114	59	16	11	16	15
BUND	10	3	32	18	21	32	38	30	91	18	39	14	16	35	29	27	21	37
DWD	31	24	17	13	10	13	45	17	50	35	47	14	19	45	23	29	31	36

Table 5: NER results DE Organisations, maximum values highlighted.

Persons	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Barack Obama	0	0	5	1616	1122	785	298	1092	622	328	313	204	364	65	45	57	25	11
Donald Trump	0	0	0	0	0	0	0	5	0	0	4	484	1332	551	301	443	142	35
Emmanuel Macron	0	0	0	0	0	0	0	0	0	4	6	3	442	382	188	216	183	250
Joe Biden	0	4	0	7	13	0	0	9	27	4	2	7	3	0	1	385	430	117
Nicolas Hulot	14	137	115	68	79	43	223	14	8	42	68	26	450	366	75	45	18	5
François Hollande	9	4	0	0	0	0	42	308	179	249	418	129	75	27	10	18	9	10
Nicolas Sarkozy	5	25	406	301	428	492	76	131	62	48	11	102	30	10	17	8	8	22
George Bush	726	348	342	419	197	130	42	54	27	7	22	33	38	3	3	10	7	2
Angela Merkel	5	8	162	78	50	36	38	50	66	48	52	23	113	62	36	70	104	13
Laurent Fabius	9	8	0	0	0	0	0	0	23	45	266	96	53	12	7	21	8	8
Vladimir Poutine	37	41	50	10	29	22	0	0	16	121	32	23	80	36	35	21	50	69
Xi Jinping	0	0	0	0	0	0	4	14	23	48	76	7	70	63	8	57	82	45
Ségolène Royal	0	8	40	3	18	36	0	14	0	176	94	148	45	15	3	1	0	1
John McCain	14	0	3	805	13	11	0	14	8	0	0	3	0	0	0	0	0	0
Hillary Clinton	0	4	8	507	42	36	4	18	58	17	14	59	13	3	1	7	2	0
Al Gore	19	166	189	71	42	33	8	5	66	14	14	10	5	12	8	3	2	3
Audrey Garric	0	0	0	0	0	0	0	0	0	0	0	0	0	12	67	57	50	51
Laurence Tubiana	5	0	3	3	61	11	13	23	8	21	67	79	23	14	20	25	10	16
Antonio Guterres	0	0	0	0	0	0	0	0	0	0	0	0	30	17	67	32	39	53
Jean Jouzel	23	33	67	3	53	22	42	14	47	17	36	10	15	21	32	15	5	13
Jacques Chirac	247	116	140	10	18	33	4	9	12	4	5	17	13	12	11	3	8	3
Greta Thunberg	0	0	0	0	0	0	0	0	0	0	0	0	0	5	150	50	25	9
Ban Ki-moon	0	0	87	24	42	7	4	18	8	97	63	53	10	0	6	3	1	1
Pascal Canfin	0	0	0	0	0	0	0	0	0	21	27	26	25	29	20	20	33	27

Table 6: NER results FR Persons, maximum values highlighted.

Organisations	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
ONU	489	153	302	294	407	449	277	290	175	349	459	399	417	199	351	260	318	260
UE	256	195	217	352	252	210	185	249	159	211	175	194	224	166	258	299	306	303
EDF	377	302	172	149	105	91	202	258	171	204	124	237	136	335	125	124	135	159
GIEC	107	104	287	105	195	499	160	82	385	335	175	59	43	168	137	75	167	150
AIE	70	157	60	122	47	58	214	59	144	142	67	59	58	67	56	81	182	149
Commission européenne	79	120	95	112	58	25	63	127	70	100	53	102	50	55	103	137	144	125
G20	0	4	0	3	231	203	88	104	74	239	96	89	221	58	54	46	110	65
Total	37	62	15	24	32	76	92	104	27	45	64	109	28	57	47	109	107	124
Verts	74	54	115	30	42	29	147	45	51	52	20	17	23	41	156	99	102	74
Banque mondiale	65	54	105	88	71	47	29	68	101	42	141	46	108	36	89	35	35	43
Greenpeace	74	21	47	68	74	29	38	54	58	35	68	26	50	70	92	73	78	48
OCDE	37	70	62	98	32	15	38	14	47	62	108	92	30	60	68	32	80	33
OMC	51	54	85	98	131	72	8	27	4	10	25	13	20	63	99	36	49	45
G7	5	4	10	14	11	15	0	18	8	10	64	10	116	86	134	21	75	46
CNRS	74	54	25	24	32	112	46	45	70	45	71	63	48	46	32	35	37	61
Sénat	65	29	17	78	95	83	25	73	43	80	49	30	35	17	17	66	65	28
CCNUCC	9	17	15	3	34	91	21	45	19	7	194	129	78	24	29	6	20	8
WWF	37	8	42	41	42	33	29	27	39	31	58	73	20	58	61	56	38	47
Parlement	33	21	13	30	39	22	8	32	23	55	31	36	45	48	67	43	63	60
FMI	14	0	10	61	108	43	13	18	35	42	84	23	35	22	54	21	22	57
Parlement européen	5	4	13	54	42	22	4	9	27	24	8	66	23	26	54	71	49	86
Conseil fédéral	163	116	157	88	13	7	21	23	8	4	7	0	0	21	49	22	47	44
G8	242	141	224	230	84	58	13	27	23	14	11	0	0	0	3	1	0	0
Shell	42	33	20	30	8	11	21	27	27	7	40	23	10	12	17	45	54	32

Table 7: NER results FR Organisations, maximum values highlighted.

Persons	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Donald Trump	0	0	0	1	0	0	2	46	2	0	23	1027	3389	1445	703	1118	277	90
Boris Johnson	4	9	11	31	21	30	8	31	42	9	51	214	181	165	1330	642	746	481
Joe Biden	0	0	1	3	6	1	1	1	1	2	6	4	7	1	31	635	1024	507
David Cameron	64	446	252	182	192	256	269	494	406	489	354	493	42	30	25	18	29	25
Barack Obama	0	0	6	236	626	219	110	200	284	475	317	310	212	68	20	57	49	16
Jeremy Corbyn	0	0	0	0	0	7	4	0	5	0	70	244	125	288	821	37	18	19
Gordon Brown	287	445	539	593	407	174	47	45	18	29	34	29	38	47	18	34	35	31
Tony Blair	1544	540	545	126	171	65	54	47	26	33	59	17	13	16	28	11	17	24
Vladimir Putin	65	66	85	19	21	24	25	9	13	197	48	77	173	42	22	34	79	330
George W. Bush	750	255	462	237	128	46	35	29	49	49	36	37	39	34	12	37	14	7
Al Gore	1	454	523	92	155	134	62	46	100	31	33	41	93	43	12	22	11	13
Liz Truss	0	0	0	0	1	0	0	1	0	18	30	31	4	1	9	11	16	561
George Osborne	1	18	21	13	14	78	215	411	276	82	131	110	22	19	16	8	7	4
Angela Merkel	26	21	116	59	58	16	27	42	19	31	60	49	293	58	50	50	71	6
Emmanuel Macron	0	0	0	0	0	0	0	0	0	0	0	5	365	220	94	58	66	57
Theresa May	0	0	0	0	1	10	4	4	7	7	11	240	286	205	153	14	10	12
Greta Thunberg	0	0	0	0	0	0	0	0	0	0	1	1	0	11	271	137	83	38
Rishi Sunak	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	34	95	259
Keir Starmer	0	0	0	0	1	0	0	12	1	0	0	3	1	31	16	72	109	167
Xi Jinping	0	0	0	0	0	0	0	0	7	47	54	31	108	24	29	54	94	60
Ed Davey	0	1	0	0	1	0	1	362	251	146	57	9	7	2	18	6	6	17
Mark Carney	0	0	0	0	0	0	0	3	2	11	33	13	30	59	126	49	38	10
Michael Gove	0	0	1	0	1	17	14	4	14	41	12	94	82	55	81	41	14	11
John Kerry	1	3	9	7	16	7	2	0	20	47	50	19	14	3	1	29	101	39

Table 8: NER results UK Persons, maximum values highlighted.

Organisations	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
EU	790	449	802	806	610	420	493	574	440	530	409	1926	1070	1214	1119	957	559	500
UN	306	200	357	330	503	436	433	346	253	402	665	504	443	404	281	367	456	414
Labour	242	260	241	192	158	280	159	157	319	208	252	211	155	408	1405	223	220	314
IPCC	58	42	173	134	160	643	150	76	427	336	75	46	42	163	67	38	93	113
Shell	108	150	87	130	104	76	107	102	96	59	274	65	106	161	90	69	118	140
BP	129	266	97	58	97	279	144	46	58	49	157	77	46	56	93	133	80	130
White House	125	39	68	57	86	33	31	39	46	103	88	72	228	91	40	114	128	88
Treasury	108	134	70	85	48	123	83	169	124	99	72	67	32	51	81	77	93	90
Congress	50	41	57	40	143	62	45	55	76	164	104	77	96	84	40	74	104	88
EPA	17	9	9	5	36	32	41	47	90	194	81	109	351	213	49	62	32	68
Amazon	37	57	29	64	68	113	57	35	29	29	32	29	42	61	169	128	126	53
Exxon	78	111	58	35	36	40	28	59	26	27	130	175	145	55	28	65	103	71
NHS	16	44	40	21	32	46	23	21	19	42	43	53	57	44	281	87	69	80
Greenpeace	71	15	23	29	140	71	13	17	32	142	57	57	70	46	30	82	93	96
Senate	6	0	0	4	188	28	32	50	20	275	24	20	181	42	43	33	78	29
G20	145	65	62	69	93	74	122	100	76	56	69	49	40	34	37	32	37	29
EDF	64	30	25	163	34	29	37	117	230	31	57	289	26	8	5	18	11	21
Met Office	36	43	46	22	44	141	72	84	185	142	67	45	41	46	14	13	9	70
Bank of England	9	18	5	21	10	13	11	8	11	20	32	20	26	57	108	84	85	82
XR	0	0	0	0	0	0	0	0	0	0	0	0	0	26	239	130	69	43
World Bank	107	55	38	111	62	54	62	97	42	39	75	44	32	29	30	33	37	29
IEA	41	27	48	70	42	23	89	75	27	56	45	35	28	26	23	33	72	60
FTSE	23	9	22	13	22	15	15	24	19	11	21	46	21	63	56	120	90	40
Nasa	16	41	21	26	53	37	41	52	41	86	65	165	97	59	8	23	17	22

Table 9: NER results UK Organisations, maximum values highlighted.

Persons	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Donald Trump	0	10	3	0	2	0	17	0	4	9	579	6200	7193	6200	5481	5699	1634	732
Joe Biden	8	0	5	54	52	23	44	23	51	8	32	18	6	29	885	3585	3832	2064
Barack Obama	0	107	291	3968	4386	3484	2102	3144	3240	2825	2737	2774	1577	650	484	707	376	181
George W. Bush	2082	979	1738	1640	771	459	290	270	167	195	255	233	183	274	97	107	55	34
Bernie Sanders	0	0	5	0	4	0	3	3	0	0	167	473	35	82	554	638	86	43
Clinton	156	101	336	900	481	368	313	187	199	188	435	1470	100	94	104	117	32	10
Scott Pruitt	0	0	0	0	0	0	0	0	0	95	8	139	740	971	33	17	7	2
McCain	103	148	112	2728	113	90	36	63	43	26	24	55	24	128	32	18	4	5
Al Gore	42	977	1015	269	194	116	125	386	201	50	76	153	56	66	24	32	15	25
Elizabeth Warren	0	0	4	13	0	0	0	5	4	12	8	27	7	45	393	175	25	13
Hillary Clinton	19	29	72	222	84	44	47	53	22	47	162	584	102	96	100	102	17	5
Vladimir Putin	38	34	62	13	17	0	17	13	12	111	123	192	188	116	71	71	76	176
Mitch McConnell	4	0	14	31	11	63	22	28	24	156	110	43	41	69	153	124	156	43
Kamala Harris	0	0	3	1	2	0	11	20	2	2	16	1	4	24	232	257	70	33
John Kerry	38	36	94	77	127	133	14	172	171	249	210	87	11	7	19	69	181	120
Xi Jinping	0	0	0	0	0	2	11	3	132	252	236	79	209	76	52	84	126	58
Manchin	0	0	3	0	0	11	3	8	10	26	3	4	9	25	10	9	287	329
Boris Johnson	69	26	54	287	23	21	6	33	12	14	16	48	25	24	118	68	140	29
Macron	0	0	0	0	0	0	0	0	0	0	0	0	225	230	106	22	31	45
Robert Mueller	0	0	0	0	0	0	0	0	0	0	0	0	21	194	228	16	3	2
Chuck Schumer	4	10	10	23	8	17	8	20	0	11	11	14	30	47	55	51	137	141
Pete Buttigieg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	208	133	52	11
Bloomberg	4	0	54	1	60	4	39	88	29	5	28	32	35	77	107	125	10	25
Mike Pence	0	0	0	0	11	4	3	10	0	3	1	165	46	62	73	171	22	10

Table 10: NER results US Persons, maximum values highlighted.

ORGANIZATION	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Senate	633	125	442	491	840	972	255	280	269	628	348	301	318	562	582	614	829	513
White House	393	171	324	434	649	821	473	371	492	520	452	420	706	739	613	520	678	355
Congress	408	268	611	573	942	954	531	396	594	616	569	497	453	543	594	383	580	516
EPA	229	257	218	661	417	392	401	490	366	680	353	258	685	928	342	351	201	382
UN	84	86	261	173	303	280	194	209	138	333	335	233	278	204	224	144	331	324
GOP	95	18	70	205	145	274	160	212	132	114	124	206	126	362	240	216	237	160
NAICS	0	0	0	0	0	0	0	0	0	0	0	0	0	476	584	178	42	32
Exxon	160	216	154	246	61	29	80	159	83	66	142	344	123	106	74	85	145	90
Supreme Court	27	36	87	47	44	55	28	58	53	76	84	219	64	158	118	178	50	206
EU	122	49	109	104	92	103	105	86	130	99	80	72	114	100	102	55	107	115
NOAA	50	75	34	19	31	25	69	53	69	31	113	55	68	111	119	155	105	56
Twitter	0	0	0	0	8	2	0	35	39	24	56	194	247	103	108	86	71	95
Medicare	0	8	15	63	61	27	42	106	47	21	63	43	33	37	190	102	99	68
Amazon	65	18	23	15	61	6	47	10	24	26	33	46	22	56	154	101	82	127
NASA	99	265	80	42	40	67	80	58	49	75	77	128	98	119	58	34	50	35
FED	8	8	20	20	20	23	36	13	35	12	68	25	13	57	83	89	125	102
Department of Energy	122	135	80	78	92	80	213	78	73	44	31	93	68	80	43	50	40	68
Democratic Party	0	8	21	44	19	48	8	35	24	35	39	60	26	46	128	119	45	27
State Department	19	5	10	15	52	34	138	88	134	150	98	58	102	75	78	42	29	20
NATO	0	3	11	23	52	6	19	8	2	8	4	116	216	85	58	53	31	21
Interior Department	11	21	9	19	20	44	25	18	18	8	34	26	57	99	54	78	58	50
Shell	31	57	23	44	12	17	89	227	92	20	192	46	16	45	23	31	77	89
FBI	0	5	0	5	4	0	3	5	6	0	18	52	48	156	93	48	10	38
Republican Party	15	13	11	47	24	31	25	48	41	40	32	68	35	74	73	57	50	36

Table 11: *NER results US Organizations, maximum values highlighted*