How People in Different Countries Talk About Climate and Energy Online

Sofia Axelsson
Sören Holmberg

UNIVERSITY OF
GOTHENBURG

Introduction

Recent technological advances have created new possibilities for systematic and quantitative text analysis in the social sciences. In terms of social media, we are now able to "listen in" on people's communication online, and use that communication – in this case, written text – to analyze public subject matters and political issues. If we take advantage of online text data, it could be considered as an unobtrusive data collection method. As an alternative, online text data can be anonymously retrieved, provided that access to data is not hampered for ethical or legal reasons. This would reduce the need for time-consuming and costly surveys.

However, there are at least two evident problems with this idea. First, the resulting data are in all likelihood much less specific than data derived from carefully designed survey questions. And second, as mentioned, collection of online data is for ethical, legal as well as practical reasons not useful for securing individual level data. Rather, it is better suited for collecting data on aggregate level, e.g. the regional or national level.

Aim and Methodology

The aim of this report is to test the extent to which online text data can be used for the study of public debate on energy and climate politics across different countries. The scope of our study is limited to nine countries, using accumulated online text data from 2015 to 2020. In this sense, it is a pilot test, not a comprehensive study. The data is retrieved from an online lexicon provided by the research project *Linguistic Explorations of Societies (LES)*, seated at the Mid Sweden University and the University of Gothenburg.[1]

The lexicon is based on a very large number of word-embedding models trained on online text data from a commercial data provider. The data consists of a vast number of web-documents from editorial (including news sources) and social (including forum sources) platforms from many different languages and countries. The lexicon can be used to retrieve information about how a given word is being used in online editorial or social media in a specific country.[2] The lexicon works as follows: one chooses the language, country and media type – each combination of the three constituting an existing word-embedding model – and types the target word into the lexicon's search field to return compiled information on how this very word is used in the chosen model.[3]

In this report, we are interested in comparing the relative frequency of certain terms pertaining to climate, environment and energy in both online editorial and social media across nine different countries. Relative frequency is calculated by dividing the occurrence of each target term in the specified model with the number of total words in that model and, for simplicity, adding a constant of 10^7.

The countries and languages we are looking into in this report is Austria (German terms), Germany (German terms), France (French terms), Denmark (Danish terms), Sweden (Swedish), Russia (Russian), the United Kingdom (English), the United States (English) and India (English). The stem words or concepts we are looking to measure are *climate, environment, emission, energy, oil, gas, nuclear power,*

---

[1] For more information about LES, see https://www.gu.se/en/linguistic-explorations-of-societies.

[2] For more information about the methodology behind the lexicon as well as the data used to build the models, see Dahlberg, Stefan, Sofia Axelsson, Amaru Cuba Gyllensten, Magnus Sahlgren, Ariel Ekgren, Sören Holmberg & Jonas Andersson Schwarz. 2022. A Distributional Semantic Lexicon for Linguistic Explorations of Societies. *Social Science Computer Review*, 41(2), 308–329.

[3] For more practical applications of the lexicon, see Dahlberg, Stefan, Sofia Axelsson & Sören Holmberg. 2020. Democracy in Context: Using a Distributional Semantic Model to Study Differences in the Usage of Democracy Across Languages and Countries. *Zeitschrift für Vergleichende Politikwissenschaft*, 14(4), 425–459.

solar power and *wind power*. However, since the lexicon can only accommodate a single word at a time, we must rely solely on stem words that pertain to the concept we are looking to measure for languages where some of these concepts consists of more than one word. An example of this is the concept *nuclear power* in English or *énergie nucléaire* in French where we can only use the stem *nuclear* or *nucléaire* when querying the lexicon for information about nuclear power. This is an evident shortcoming of our methodology, given that a search for the word nuclear is a lot broader than simply nuclear power and thus covers all searches related to something nuclear, whereas a search for nuclear power in Danish, German or Swedish will be much more specific since each translation of the concept consists of only one word (e.g. *atomkraft* in Danish, *Kernkraft* in German, and *kärnkraft* in Swedish). To make up for this, we have tried to include an additional number of common words pertaining to nuclear power where applicable (e.g. nuclear energy, nuclear power plant, nuclear disaster, etc.).

Because we are dealing with languages that come from different language families – i.e. Germanic, Romance, Slavic – and therefore has different structures, we have consulted with native speakers of each language to make sure that we not only include the most correct translations of the concepts we are looking for, but that we also include relevant inflections of the stem words or stem concept we are looking to measure. In some languages two translations of the same concept may be equally used (e.g. *Kernenergie* and *Kernkraft* in German).

Results and Discussion

Comparing the terms related to energy, environment and climate in editorial media and social media in the nine chosen countries, our expectation is that the relative frequency of terms should be higher in editorial media than in social media, given that the former is more "elitist" and the latter "folksier". The results are presented in Table 1 with relative frequencies and in Table 2 with standardized quotients and prove our expectation right in a clear majority of cases. Mentions of the selected terms are more frequent in editorial media than in social media in 60 out of a total of 90 cases (67 percent). More specifically, we have 60 cases where the term mentions are more frequent in editorial media, 6 cases where the term mentions are equally frequent in editorial as in social media, and 24 cases where the term mentions are most frequent in social media.

Our expectation that the relative frequency is higher in editorial media is most clearly supported for terms such as climate, emission, nuclear and gas. Two terms do not fit our expectations. These are energy and oil. With regard to these terms, more countries (five countries for energy and six countries for oil) reveal higher frequencies in social media than in editorial media. We have no evident explanation for why this is. Neither can we explain why our expectation fits better in some countries than others. A perfect fit to the expectation is found in the US where all terms selected and measured yield a higher relative frequency in editorial than in social media. In Denmark and France, 8 out of 10 terms yield higher frequency in editorial media. Least support for the expectation is found in India where a clear majority of terms yielded a higher frequency in social media rather than editorial media, and in Austria where half of the terms rendered a higher frequency in editorial media, and the other half rendered a higher frequency in social media. A (perhaps speculative) explanation for the divergent outcome in India is the fact that the language of analysis is English. We might therefore suspect that English social media reflect a more "elitist" and higher-educated segment of the Indian population.

From a more normative democratic perspective, similarities in mentions across editorial and social media could be considered a positive outcome as it suggests that more elitist communication is not that different from popular communication. Put differently, elite communication should, from this perspective, more closely represent the communication of ordinary people. Approaching the results in

this way – and disregarding the special case of India – the countries in our study are ranked from best (in terms of least observed differences between the two media types) to worst in the following order: 1) Austria, 2) the United Kingdom, 3) Russia, 4) Germany, 5) France, 6) Sweden, 7) the United States, and 8) Denmark. The country with the largest difference between editorial and social media in terms of relative term frequency is Denmark. The specific reason for this is that the terms selected to measure wind power yield a very high relative frequency compared to other countries; wind power is evidently a topic of much conversation in Danish online editorial media.

However, if we turn our attention to the relative term frequencies in standardized form, it is clear that the divergencies between editorial media and social media are very modest. Two terms reveal the smallest differences: energy and solar/solar power. Mentions of the term energy have the highest frequency across all countries. For solar/solar power, the opposite is true; mentions of this term have the lowest frequency across all countries, with the possible exception of India. The largest difference across countries can be observed for the term wind/wind power. Here, Denmark represents an outlier with a term frequency that is two or three times higher than all the others countries' term frequency. The lowest frequency of the term wind/wind power is found in Russia and the US.

Comparing the standardized term frequencies, one observation stands out. Terms related to fossil fuels are placed as the number one term with highest frequency in all countries. The term oil was ranked as number one in editorial online media in Denmark, Sweden, the UK, the US and India, whereas gas was ranked as number one in Austria, Germany, France and, especially, Russia. In social media, oil is further ranked as number one in Austria, Germany, Denmark, Sweden, the UK, the US and India. Gas is the top term in social media for Russia and France. Thus, oil is more frequently communicated about amongst the "ordinary people" than amongst the "elite", at least measured this way.

Concluding Remarks

A conclusion from this very modest test using online text data from the LES project, is that it is useful to pay attention to what people communicate online. The results are not revolutionary, but they give some new insight. The "elite-mass" comparisons and the differences between countries as well as between divergent terms related to climate and energy politics are worth following up in future studies. Our preliminary results indicate the usefulness of online text data and natural language processing is worth pursuing for climate and energy-related studies in the social sciences.

Table 1. Relative Frequency of Climate and Energy Terms in Editorial Media and Social Media

	Sweden			Denmark			Germany		
	Editorial	Social	Diff	Editorial	Social	Diff	Editorial	Social	Siff
Climate	68	46	+22	96	38	+58	42	36	+6
Environment	109	85	+24	145	70	+75	51	40	+11
Emission	46	24	+22	19	8	+11	17	15	+2
Energy	85	136	-51	119	113	+6	72	76	-4
Nuclear (power)	15	9	+6	1	1	±0	7	5	+2
Coal	11	11	±0	10	7	+3	11	19	-8
Oil	36	66	-30	50	88	-38	30	54	-24
Solar (power)	10	9	+1	11	7	+4	6	5	±0
Wind (power)	11	7	+4	26	7	+19	7	4	+3
Gas	16	14	+2	31	23	+8	41	35	+6

	Austria			France			Russia		
	Editorial	Social	Diff	Editorial	Social	Diff	Editorial	Social	Siff
Climate	66	65	+1	44	31	+13	37	25	+12
Environment	58	65	-7	177	80	+97	68	65	+3
Emission	22	19	+3	87	62	+25	18	11	+7
Energy	90	108	-18	101	85	+16	13	17	-4
Nuclear (power)	8	7	+1	57	40	+17	12	8	+4
Coal	10	14	-4	11	11	±0	8	7	+1
Oil	33	40	-7	41	25	+16	55	35	+20
Solar (power)	6	5	+1	21	24	-3	16	16	±0
Wind (power)	10	15	-5	7	6	+1	0	0	±0
Gas	39	30	+9	90	48	+42	148	115	+33

	United Kingdom			United States			India		
	Editorial	Social	Diff	Editorial	Social	Diff	Editorial	Social	Siff
Climate	95	72	+23	76	52	+24	64	74	-10
Environment	198	126	+72	182	107	+75	165	166	-1
Emission	44	32	+12	29	16	+13	29	30	-1
Energy	166	161	+5	209	146	+63	181	201	-20
Nuclear (power)	51	37	+14	75	37	+38	69	55	+14
Coal	15	13	+2	24	12	+12	54	31	+13
Oil	111	133	-22	173	143	+30	111	177	-66
Solar (power)	23	31	-8	45	35	+10	66	75	-9
Wind (power)	46	51	-5	50	48	+2	32	35	-3
Gas	96	73	+23	130	68	+62	99	118	-19

Comment: Relative frequency is calculated by dividing the occurrence of each target term in a given model with the number of total terms that exists in that model, and then adding a constant of 10^7.

Source: LES Distributional Semantic Lexicon (Dahlberg et al. 2022).

Table 2. Standardized Quotients of Mentions of Climate and Energy Terms in Editorial Media and Social Media

Term	Sweden	Denmark	Germany	Austria	France	Russia	UK	US	India	Mean
Climate	1,5	2,5	1,2	1,0	1,4	1,5	1,3	1,5	**1,2**	1,5
Environment	1,3	2,1	1,3	**1,1**	2,2	1,0	1,6	1,7	**1,0**	1,5
Emission	1,9	2,4	1,1	1,2	1,4	1,6	1,4	1,8	**1,0**	1,5
Energy	**1,6**	1,1	**1,1**	**1,2**	1,2	**1,2**	1,0	1,4	**1,1**	1,2
Nuclear (power)	1,7	1,0	1,4	1,1	1,4	1,5	1,4	2,0	1,3	1,4
Coal	1,0	1,4	**1,7**	**1,4**	1,0	1,1	1,2	2,0	1,7	1,4
Oil	**1,8**	**1,8**	**1,8**	**1,2**	1,6	1,6	**1,2**	1,2	**1,6**	1,5
Solar (power)	1,1	1,6	1,0	1,2	**1,1**	1,0	**1,3**	1,3	**1,1**	1,2
Wind (power)	1,6	3,7	1,8	**1,5**	1,2	1,0	**1,1**	1,0	**1,1**	1,6
Gas	1,1	1,3	1,2	1,3	1,9	1,3	1,3	1,9	**1,2**	1,4
Mean	1,5	1,9	1,4	1,2	1,4	1,3	1,3	1,6	1,2	–

Comment: Standardized quotients are calculated by dividing the relative frequency of a given term in editorial media by the relative frequency of that same term in social media, except for cases where the relative frequency is higher in social media than in editorial media. In these cases, marked in bold in the table, the relative term frequency in social media has, instead, been divided with the relative term frequency in editorial media, thereby making the frequencies comparable across different countries.

Source: LES Distributional Semantic Lexicon (Dahlberg et al. 2022).

Swedish Opinions on Environment, Energy and Climate Changes (EECC) is a research project at the Department of Political Science at the University of Gothenburg. EECC contributes to the knowledge supply of Swedish public opinion on issues related to energy, environmental and climate politics.

Linguistic Explorations of Societies (LES) is an interdisciplinary research project at the Department of Political Science at the University of Gothenburg and the Department of Humanities and Social Scientist at the Mid Sweden University. LES utilizes language technology and online text data to address the challenges facing the changing landscape of comparative public opinion and survey research.