Measuring the Sociolinguistic Patterns of Climate Debate Polarization in the Facebook Context

Luigi Arminio[1] and Luca Rossi[2]

[1] IT University of Copenhagen, Copenhagen 2300, Denmark
luia@itu.dk
[2] IT University of Copenhagen, Copenhagen 2300, Denmark
lucr@itu.dk

Abstract. This research focuses on the sociolinguistic patterns characterizing the polarized climate change debate on Facebook, focusing on the communication dynamics occurring within pro-climate action and anti-climate action stakeholders. Our study specifically aims to (1) identify the variations in language codes among these groups, and (2) assess how these linguistic nuances affect the respective audiences. For this goal, we compiled a comprehensive list of relevant English-speaking stakeholders in the climate debate and collected over 2000 of their posts spanning several months. To analyse the textual content that they produced, we defined a series of quantitative language code indicators, measuring the readability, concreteness, subjectivity and scientificity of the language used, alongside topic modeling to dissect the discussions' themes. Furthermore, we applied regression modeling to assess the impact of language code variations on the audience responses of the two debate groups. The results revealed significant variations in audience reactions across the debate spectrum, with the pro-climate audience responding more to variations in language style, whereas the anti-climate audience exhibited a distinct response to shifts in topic focus.

Keywords: sociolinguistics, climate, Facebook, polarization, engagement

1 Introduction

Over the years, various social researchers examined the interplay between social groups and language practices, with a specific focus on how socio-economic stratification affects communication styles in some well-documented studies ([1,2,3]. More specifically, drawing from the existing social science literature, it has been documented that higher socioeconomic status (SES) facilitates the development of advanced language abilities [1,2,3]. The studied nexus between SES (Socio-Economic Status) and the development of linguistic abilities prompts wider considerations about its broader implications in various sectors of society, including the issue of political polarization. In effect, beyond their documented effect on language codes, societal and economic circumstances impacting individuals have been observed to shape political positions in polarized discussions, such as the

Table 1. Data sources for the climate stakeholders list

Sources
Climate Discourse Research [15]
International Rescue Committee [16]
Business Insider [17]
Classy.org [18]
Food Tank [19]
Apolitical.co [20]
Desmog Climate Disinformation Database [21]
Feedspot [22] [23]

debate surrounding climate debate [4]. Considering this interplay alongside the association between SES and LC (language code), it suggests the plausible hypothesis that divergent language patterns may characterize the opposing sides of the debate, potentially intensifying the already existing polarization and intensifying social divides.
To test this hypothesis, we compiled a list of climate debate stakeholders (both representing pro and anti climate action stances) drawing from a combination of relevant articles and databases and gathered their Facebook posts spanning more than 11 months. After that, drawing upon a comprehensive collection of sociolinguistic literature and computational linguistic tools, we proceeded to define and establish a set of quantitative indicators to determine the LC characterizing the textual content of each post. These indicators are intended to quantify the readability, concreteness, subjectivity, and relative frequency of scientific terminology in the posts. Alongside these structural linguistic features, in light of the notoriously multifaceted nature of the climate change issue [5], we applied topic modeling across the entire corpus of posts to assign a specific thematic label to each post. Following this, we categorized each stakeholder based on the average values of the previously computed language indicators in their posts, and we measured the incidence of the different topics of discussion across the opposing sides of the climate debate.

After that, we used the gathered data to answer four research questions:

RQ1) Are there discrepancies in the language code of the two opposing sides of the polarized climate action debate?

RQ2) What are the main themes of discussions within the opposing sides of the polarized climate action debate?

RQ3) Do disparities occur in the frequency of topics discussed between the opposing sides of the polarized climate debate?

RQ4) How do the specific sociolinguistic patterns of opposing sides in the polarized climate action debate influence audience responses?

2 State of the Art

The study of language choices and practices as a socially structured phenomenon has periodically attracted interest within social science research [6,7]. In this area of study, Bernstein [7] [8] connected social stratification and language production, outlining two sociolinguistic codes: a restricted code, associated with the working class and characterized by limited syntax and context-dependent communication, and an elaborated code, linked with the middle/high class and marked by a wider vocabulary, complex sentence structures and formal, professional communication style [7] [8]. The sociolinguistic perspective brought by Bernstein on the relationship between social structures and LC usage is fundamentally based on the way in which the SES of the individuals affect their linguistic production. However, while Bernstein's theoretical contribution is crucial, language production of individuals may also be associated with other factors beyond SES alone. From this angle, additional theoretical literature suggests the potential connection between the political landscape and the communicational patterns of the individuals, theorizing a relationship between the adherence to a certain set of features of a given language code and the obtained political success [9]. In connection with this, it is notable that the diversity of communication styles in political environments has been empirically analysed, with studies drawing on political party manifestos [10] and interviews with politicians [11], revealing both the restricted accessibility of certain parties' communication styles [10] and the enhanced accessibility of language codes (in this context, meaning higher readability and simplicity) employed by specific political figures [11].

Reflecting on the posited and discussed relevance of communication patterns in political communication, it must be noted that although some studies discuss language codes' impact on political discourse, there's a lack of systematic examination and measurement of the full range of linguistic features determining the stylistic diversity of social actors and how they impact political polarization dynamics, especially in social media contexts characterized by a significant level polarization of the debates, such as the climate change debate [12]. Addressing this gap can facilitate the assessment of the potential impact of sociolinguistic patterns on polarization in public debates and may support the development of communication and divulgation approaches that facilitate constructive engagement and bridge ideological divides.

To address this gap, it is essential to define quantitative measures to identify the LC features exhibited by social media users. In this regard, the existing literature in computational linguistics offers a variety of metrics to evaluate the readability of textual documents [24]. These formulas are aimed at assessing the accessibility and the understandability of a text through parameters such as sentence length, word length, and syllable count [24], which also appear in Bernstein's empirical exploration of LC and speech patterns[7]. However, although these metrics represent a crucial contribution to assess the morphosyntactic complexity of textual communication, they primarily focus on surface-level attributes (such as the above-mentioned sentence length and syllable count). While these metrics are useful for assessing text complexity, they don't capture the semantic variations, which are crucial for a comprehensive evaluation of language codes. For example, a social media post characterized by simple sentence structures may still express sophisticated concepts through the resort to a technical or specialized lexicon. In this sense, to provide a comprehensive evaluation of the language style of the individuals, it is crucial to consider the broader spectrum of semantic and contextual features that characterize the messages. Consequently, we found it essential to integrate these measures with lexicon-based metrics for capturing the intricate semantic nuances and contextual features that define communication.

Building upon this premise, this study establishes and selects a series of quantitative indicators to measure the sociolinguistic and communicative patterns prevalent among stakeholders in the highly polarized climate change debate on social media and the impact of communicative and stylistic variations on their audience.

3 Data and Methods

To pursue our research, we monitored the Facebook activity of a cohort of English-speaking stakeholders in climate debate. The list of stakeholders, manually compiled, includes actors known for pro-climate action positions (*pro-actors*, or *pro-climate*) and opponents of climate action stances (*counter-actors*, or *anti-climate*). This binary representation (pro-anti climate) is based on the well-documented strong polarization of the social media debate on climate change [12,**?**]. To compile the list, we resorted to multiple sources, including scholarly research on climate discourse, content aggregation platforms and databases from NPOs (Non-profit organizations) and policy insight networks (TABLE 1). Our final list of actors encompasses, for pro and counter actors, journalists, NPOs, for-profit organizations, academicians, political figures, and other prominent voices in the climate action debate. Only those actors who displayed recent Facebook activity, defined as having posted at least once in 2023 or 2024, were catalogued.

After defining the list, we resorted to the Crowdtangle API to collect textual data from these actors over 11 months, from March 1, 2024, to February 12, 2024, limiting to 100 posts per actor. We collected 10,650 posts from 250 actors, including 96 pro-climate and 156 anti-climate actors. We then filtered posts for climate relevance based on a dictionary of climate-related terms[14], resulting in 2,326 relevant posts from 188 actors. For each of these posts, we calculated various LC indicators (i.e., a readability score, a subjectivity score, a concreteness score, and a scientificity score).
First, to measure users' LC from a morphosyntactic standpoint, we examined the existing literature in computational linguistics and identified the most recognised readability scores, commonly employed in NLP to assess text comprehensibility [24] [28] . Then, we assessed the internal consistency of the selected readability metrics by computing Cronbach's alpha [25] of these measures on the entire corpus, aiming for values above the conventional 0.7 [26] [27] threshold. This step was essential to verify that the readability indexes were consistently measuring

the same underlying concept of text comprehensibility. Given the satisfactory result of 0.78, we averaged all the selected readability indices mentioned into a single average readability measure AR:

$$\mathrm{AR}_i = \frac{1}{8}\sum_{j\in J}\mathrm{RD}_{ij} \tag{1}$$

where J is the set of measures {Kincaid, ARI, Coleman-Liau, FleschReadingEase, GunningFogIndex, LIX, SMOGIndex, RIX}[28] and RD_{ij} is the z-standardized score of document i according to index j.

Following the establishment of a structural readability measure, we developed a quantitative semantic index focused on the concreteness-abstractness of the utilized lexicon, which is also a recurring theme in Bernstein's examination of language production variations [7]. For this study, we defined concreteness of a textual document as the stylistic inclination towards lexical choices that that invoke perceptible entities [29]. First, to define the concreteness level of the words, we used a database assembled from an extensive study that compiled concreteness ratings for more than 37,000 English lemma words [29], each of them rated from 1 to 5 based on their concreteness level. To enhance computational efficiency and interpretability, we conducted a min-max scaling on these ratings, placing word lemmas on a continuous concreteness scale from 0 to 1.

Then, we lemmatized our posts for consistent matching with the concreteness ratings database, we computed the concreteness score of each post's lexicon. To do this, we retrieved the concreteness score for each lemma from the ratings database and weighted it by its TF-IDF (term frequency–inverse document frequency) value within the document to highlight significant words. We then combined the TF-IDF weighted concreteness scores for each lemma word of the post, dividing the total by the sum of the TF-IDFs of each lemma, representing the highest concreteness scenario. This resulted in a normalized concreteness score, denoted as WAC (Weighted Average Concreteness), ranging from 0 to 1:

$$\mathrm{WAC}_i = \frac{\sum_{k\in K}(\mathrm{NC}_{ik}\times \mathrm{TF\text{-}IDF}_{ik})}{\sum_{k\in K}\mathrm{TF\text{-}IDF}_{ik}} \tag{2}$$

where K is the set of lemmas in post i that also appear in the concreteness ratings database, NC_{ik} is the min-max normalized concreteness score of lemma k in document i, and $\mathrm{TF\text{-}IDF}_{ik}$ is the TF-IDF weight of lemma k in document i.

After that, through the TextBlob Python library, we labeled our posts according to their subjectivity score, which measures the prevalence of perspectival influence over objective neutrality [30]. In the subjectivity assessment of any document D, TextBlob assesses subjectivity based on its constituent words $w_1, \ldots, w_n$ using a database containing English adjectives. Each adjective a in this database is assigned a subjectivity level within the range of 0 to 1 [31]. We opted to consider subjectivity as a pertinent language code indicator because subjective communication can be considered indicative of an informal LC [32].

Subsequently, we defined a metric to assess the prevalence of scientific language within the posts, motivated by the known association between the resort to scientific jargon and restricted language accessibility [33]. For this goal, we adopted the LSdDC dictionary of scientific stem words, derived from a large corpus of scientific abstracts [34], encompassing over 100000 stems frequently encountered in scientific literature[34]. To uniform the matching with the LSdDC stem words, we stemmed words within posts. We assessed stemmed words for scientific relevance using a dummy variable, SCI_{ij}, where 1 indicated presence in the LSdDC dictionary and 0 otherwise for any stemmed word j in a post i. Each term's scientific significance was then weighted by its TF-IDF score, mirroring the methodology of the WAC. This process culminated in an average scientificity score for each post, denoted as Weighted Average Scientificity (WAS), described by this formula:

$$\mathrm{WAS}_i = \frac{\sum_{j \in J} (\mathrm{SCI}_{ij} \times \text{TF-IDF}_{ij})}{\sum_{j \in J} \text{TF-IDF}_{ij}} \tag{3}$$

Here, J is the set of all stems in document i, SCI_{ij} is the above-mentioned binary indicator of the scientific relevance of stem j, while TF-IDF_{ij} is the TF-IDF weighting of stem j in document i.

After establishing the LC indicators, we examined how variations in the LC impact the audience response in both pro-climate and anti-climate actors. For this goal, we used 3 audience response metrics: the Engagement Rate (ER) the Emotional Polarity score (EP), and the Emotional Diversity (D) score.
The ER, used to assess audience response on Facebook[35], lacks a universally recognized calculation method, as recognized by Facebook itself [35]. In this research, we adopt the one defining ER as the extent of interaction generated by a post given the author's audience size [35]. Within this approach, the engagement rate ER of a post p by an author a can be expressed as:

$$ER_{p_a} = \frac{I_{p_a}}{A_a} \tag{4}$$

where I_{p_a} is the sum of likes, shares, comments, and emoji reactions received by post p from author a, and A_a is the audience size (i.e., the subscriber count of the Facebook page belonging to a).

For the measure the direction of the emotional response of the audience, we resorted to the Emotional Polarization score [36], EP, that computes the direction of the emotional response of the audience to a given post as:

$$EP = \frac{n_{\text{angry}}}{n_{\text{love}}} + 1 \tag{5}$$

where n_{angry} and n_{love} are the numbers of angry reactions and love reactions received by a given post [36].

To calculate the audience response diversity, we resorted to the emotional diversity score [37]. This score evaluates the emoji-based reaction dispersion on

Facebook posts via Jensen-Shannon distance (JSD). Emotional diversity D is calculated as

$$D = 1 - \mathrm{JSD}(p, q) \tag{6}$$

where p is the normalized emoji-reaction distribution and q is a uniform distribution. This score D ranges from 0 (concentration on one emoji) to 1 (even spread) [37].

After establishing the LC indicators, we labeled each post according to its main topic using BERTopic [39]. This comprehensive methodology enhances our analysis by integrating structural linguistic variations with thematic content, moving beyond a mere focus on LC fluctuations to include the content of the communication in the exploration and the anlysis of the sociolinguistic patterns. This approach prevents the oversimplification that could result from focusing solely on linguistic structures while overlooking the content, which may also interact with language style variations to define distinct language patterns [40]: this is particularly relevant in reportedly multifaceted issues, such as the climate change debate [41].

To select the optimal number of topics, we resorted to the Calinski-Harabasz (CH) score [38] on the embedding scores of the posts of each topic, generated by the pre-trained sentence-transformers model All-MiniLM-L6-v2 [42]. In this context, topics were treated as clusters of embedding vectors corresponding to the posts, and the CH score was employed to assess the balance between intra-topic cohesion and inter-topic separation across models with different numbers of topics. We calculated the CH score CH_i for each model m_i ranging from m_2 to m_{10}, where i represents the topic number. Subsequently, we selected the model with the highest CH.

At this point, to answer the first research question (RQ1), we calculated the Point Biserial correlation between the position of each climate debate actor (pro vs counter) and their average scores in the different LC indicators. As for RQ2 and RQ3, we addressed them through the exploration of the main topics identified through BERTopic (RQ2) and the examination of the incidence of each found topic among the two opposing sides of the debate (RQ3). To answer RQ4, we fitted, for each audience response metric, three multiple linear regression models (one performed on the subset of Facebook data produced by the *pro-actors*, and one on the analogue *counter-actors* subset of posts) where we examined how variations in the different LC indicators, designated as predictors, including the topic label as a control variable among them, impacted on each of the three audience response metrics, assigned as dependent variables of the models.

Table 2. Correlation between language code indicators and climate action stances

Language Code Indicator	PB Correlation	*p*
Readability Score (AR)	0.18	0.01
Concreteness (WAC)	0.17	0.02
Subjectivity (SJ)	-0.09	0.23
Scientificity (WAS)	-0.09	0.20

Point biserial correlations between language code indicators and climate action stances. Positive values indicate a positive association with pro-climate stances, whereas negative values indicate an association with anti-climate stances.

4 Results

4.1 Correlation Analysis

We conducted a correlation analysis to examine the relationship between the LC employed by the climate debate actors and their stance on the climate change issue (pro-climate vs anti-climate).

The correlation analysis, described in Table 2, illustrates the relationships between the LC markers and the audience response metrics. A slight positive correlation with pro-climate stances was found for AR (0.18) and WAC (0.17) in the language style, suggesting that clear and tangible language may be more characteristic of the pro-climate actors, even though the relationship is modest. In contrast, subjectivity and scientificity in the language showed small negative correlation values with pro-climate stances, though the exceedingly low coefficient (-0.09) suggest that their relationship is negligible. Albeit not very strongly, it is the pro-climate actors who show tendencies towards a less elaborated LC, particularly with texts that are more accessible in terms of readability and resort to a more concrete vocabulary. This reflects an unexpected trend, given that a less elaborated LC is commonly associated with lower socioeconomic backgrounds [7], which reportedly exhibited reduced pro-environmental stance [4]. However, it must be noted that this correlation was found using social media data released by a selection of actors including academicians, politicians and business representatives from upper-class backgrounds. While it sheds light on the communication styles of key players in the climate debate, it may not necessarily represent the language pattern variations in the broader population of climate action proponents and antagonists.

4.2 Topic Analysis

Using the CH Score [38], we identified three as the optimal number of topics, that were categorized as: Climate Change Action and Policies ($CLMCH\_ACT$), Sustainable Development Practices ($SUST\_DEV$), and Political Commentary and Media Regulation (POL_COMM)

In both groups, discussions about sustainable development appear to be prominent, yet it's noteworthy how a subject that is largely missing among the

Table 3. Topics and Focus Areas

Topic	Focus (Description)	Incidence
CLMCH_ACT	Mitigation, emission reduction advocacy, fossil fuel exit	pro: 34.12% counter: 29.82%
SUST_DEV	Sustainable agriculture, renewable energy, environmental conservation	pro: 65.81% counter: 44.07%
POL_COMM	Influence on public opinion, freedom of speech, censorship	pro: 0.07% counter: 26.11%

pro-climate actors is considerably addressed by counter-actors, namely, political discussions with a significant emphasis on censorship.

4.3 Regression Analysis

Regression analysis was employed on the two sub-corpora of posts from pro and counter actors to assess the influence of the LC indicators on their audience reactions, with the inclusion of the topic as a control variable.

The results, visible in TABLE 4, showed significantly different patterns in the two groups: the audience of pro-climate actors was notably more influenced by LC variations in the posts compared to the one of counter-actors. Conversely, the counter-actors audience was more responsive to topic variations, a pattern not observed in the pro-climate audience. Additionally, when a parameter affected the same LC marker for both groups, it exhibited opposite effects more than once.

The clearest difference between the two groups is seen in the readability of the posts, which significantly differs in its impact on all the considered audience response dimensions across the two groups: for the pro-actors audience, it increases engagement and fosters positive over negative reactions with reduced emotional diversity, while for the counter-actors, it is irrelevant for both engagement and diversity and shifts the emotional tone of reactions toward negativity.

In both groups, more frequent resort to the scientific lexicon appears to make the emotional response of the audience more negative. This may relate to a perceived complexity of the scientific language, which could foster a sense of disconnect from the audience. However, more research is needed to fully explore this phenomenon.

5 Discussion and Conclusion

Our study revealed distinct sociolinguistic dynamics characterizing the two sides of the polarized climate change debate on Facebook. While the two groups did not show dramatic divergencies in their language style, a higher presence of readable and concrete language was noted among pro-climate actors. As for the main topics of interest, anti-climate actors exhibited a distinct focus on

Table 4. Regression Results for PRO and COUNTER actors

	PRO			COUNTER		
Metric	Estimate	p	R-squared	Estimate	p	R-squared
Engagement Rate						
Intercept	0.008	0.021*	0.011	-0.061	0.348	0.012
Avg. Readability	0.002	0.001**		0.0001	0.990	
Concreteness	-0.002	0.623		0.027	0.733	
Scientificity	-0.003	0.371		0.096	0.119	
Subjectivity	0.001	0.616		-0.004	0.890	
CLMCH ACT	0.001	0.303		0.041	0.003*	
POL COMM				0.007	0.623	
Emotional Polarization						
Intercept	0.354	0.002**	0.030	0.424	0.003*	0.041
Avg. Readability	-0.088	<0.001***		0.046	0.004*	
Concreteness	0.177	0.252		-0.002	0.990	
Scientificity	0.479	<0.001***		0.661	<0.001***	
Subjectivity	0.043	0.430		-0.026	0.698	
CLMCH ACT	-0.019	0.353		-0.006	0.838	
POL COMM				0.031	0.328	
Emotional Diversity						
Intercept	0.332	<0.001***	0.068	0.438	<0.001***	0.094
Avg. Readability	-0.041	<0.001***		0.003	0.405	
Concreteness	0.005	0.900		0.00004	0.999	
Scientificity	0.121	<0.001***		0.028	0.396	
Subjectivity	0.028	0.031*		-0.040	0.017*	
CLMCH ACT	-0.006	0.199		-0.010	0.199	
POL COMM				0.062	<0.001***	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

discussions that frame climate debate in a strongly political dimension with a particular focus on the theme of free speech and censorship. The focus on issues related to freedom of speech and mainstream media among anti-climate actors reflects a broader existing narrative that highlights resistance to perceived mainstream media censorship and raises concerns over elite dominance in public discourse[43]: this narrative is common in right-leaning circles[43], where much anti-climate rhetoric finds resonance[44].

In terms of user reactions, a significant divergence in the factors affecting audience response to posts from pro-climate or anti-climate actors was detected: while the pro-climate action audience showed higher sensitivity to variations in language code, the anti-climate audience was responsive to shifts in topic focus. Furthermore, in both groups, a higher presence of scientific terms triggered more negative reactions, suggesting that excessively technical language could be a barrier in climate-related communication.

Overall, these findings highlight the need for tailored communication strategies that consider language codes and topic focus to effectively interact with diverse audiences in order to bridge divides and foster constructive dialogue concerning the climate crisis.

References

1. Schneider, T., & Linberg, T. (2022). Development of socio-economic gaps in children's language skills in Germany. *Longitudinal and Life Course Studies*, *13*(1), 87-120.
2. Weiland, C., & Yoshikawa, H. (2014). Does higher peer socio-economic status predict children's language and executive function skills gains in prekindergarten? *Journal of Applied Developmental Psychology*, *35*(5), 422-432.
3. Hawkins, P. R. (1969). Social class, the nominal group and reference. *Language and Speech*, *12*(2), 125-135.
4. Weckroth, M., & Ala-Mantila, S. (2022). Socioeconomic geography of climate change views in Europe. *Global Environmental Change*, 72, 102453.
5. de Guttry, C., Süsser, D., & Döring, M. (2019). *Situating climate change: Psychological distances as tool to understand the multifaceted dimensions of climate change meanings*. Geoforum, 104, 92–100.
6. Ochs, E., & Schieffelin, B. B. (2011). The theory of language socialization. The handbook of language socialization, 1-21.
7. Bernstein, B. (1962). Linguistic codes, hesitation phenomena and intelligence. Language and speech, 5(1), 31-48.
8. Bernstein, B. (1964). Elaborated and restricted codes: Their social origins and some consequences. American anthropologist, 66(6), 55-69.
9. Ufaeva, V. (2013). Linguo-Communicative Code of Modern Political Interaction. Historical, Philosophical, Political and Law Sciences, Culturology and Study of Art. Issues of Theory and Practice, 6(32), 192-196.
10. Gyasi, W. K. (2023). The Readability of Political Party Manifestos of the 2016 General Elections in Ghana. Athens Journal of Mass Media & Communications, 9(1).
11. Kayam, O. (2018). The readability and simplicity of Donald Trump's language. Political Studies Review, 16(1), 73-88.

12. Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., ... & Baronchelli, A. (2022). Growing polarization around climate change on social media. Nature Climate Change, 12(12), 1114-1121.
13. Zhou, J. (2016). Boomerangs versus javelins: How polarization constrains communication on climate change. Environmental politics, 25(5), 788-811.
14. United States Environmental Protection Agency, "Glossary of Climate Change Terms," January 19, 2017. [Online]. Retrieved from 19january2017snapshot.epa.gov [Accessed: February 20, 2024].
15. Chen, K., Molder, A. L., Duan, Z., Boulianne, S., Eckart, C., Mallari, P., & Yang, D. (2023). How climate movement actors and news media frame climate change and strike: evidence from analyzing twitter and news media discourse from 2018 to 2021. *The International Journal of Press/Politics*, 28(2), 384-413.
16. International Rescue Committee. (2024). *12 climate activists inspiring us to fight climate change*. Last consulted on March 4, 2024. Retrieved from rescue.org
17. Business Insider. (2023). *30 global leaders climate solutions*. Last consulted on March 4, 2024. Retrieved from businessinsider.com
18. Classy.org. (2024). *Climate change organizations*. Last consulted on March 4, 2024. Retrieved from classy.org
19. Food Tank. (2020). *36 organizations helping solve the climate crisis*. Last consulted on March 4, 2024. Retrieved from foodtank.com
20. Apolitical.co. (2022). *Most influential people in climate policy for 2022*. Last consulted on March 4, 2024. Retrieved from apolitical.co
21. Desmog. (2024). *Climate Disinformation Database*. Last consulted on March 4, 2024. Available: https://www.desmog.com/climate-disinformation-database/
22. Feedspot. (2024). *Top 45 Climate Change Activists Influencers in 2024* . Last consulted on February 20, 2024. Retrieved from influencers.feedspot.com
23. Feedspot. (2024). *Top 100 Sustainability Influencers in 2024* . Last consulted on February 20, 2024. Retrieved from influencers.feedspot.com
24. Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. Reading in a foreign language, 23(1), 84-101.
25. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297-334.
26. Adeniran, A. O. (2019). Application of Likert scale's type and Cronbach's alpha analysis in an airport perception study. Scholar Journal of Applied Sciences and Research, 2(4), 1-5.
27. Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in science education, 48, 1273-1296.
28. Van Cranenburgh, A. (2019). *readability* (Version 0.3.1) [Software]. PyPI. Retrieved from pypi.org
29. Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior research methods, 46, 904-911.
30. Verma, P. (2022). Sentiment Analysis Using Subjectivity and Sentiment Score. *The Next Normal: Building Agile, Sustainable, Tech-enabled Organizations (TNN–BASTO-2022)*, 55.
31. Steven Loria. TextBlob. 2024. `github.com/sloria/TextBlob`. [Computer software].
32. Chute, A., Johnston, S., & Pawliuk, B. (2023). *Professional Communication Skills for Health Studies*.

33. Bullock, O. M., Colón Amill, D., Shulman, H. C., & Dixon, G. N. (2019). Jargon as a barrier to effective science communication: Evidence from metacognition. *Public Understanding of Science*, **28**(7), 845-853.
34. Suzen, N., Mirkes, E. M., & Gorban, A. N. (2019). LScDC-new large scientific dictionary. *arXiv preprint* arXiv:1912.06858.
35. Chorana, A., & Cherroun, H. (2021). User Generated Content and Engagement Analysis in Social Media case of Algerian Brands. In Proceedings of the 4th International Conference on Natural Language and Speech Processing (ICNLSP 2021) (pp. 156-165).
36. Muraoka, T., Montgomery, J., Lucas, C., & Tavits, M. (2021). Love and anger in global party politics: Facebook reactions to political party posts in 79 democracies. Journal of Quantitative Description: Digital Media, 1.
37. Freeman, C., Alhoori, H., & Shahzad, M. (2020). Measuring the diversity of Facebook reactions to research. Proceedings of the ACM on Human-Computer Interaction, 4(GROUP), 1-17.
38. Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1-27.
39. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
40. Cartright, M. A., Aktolga, E., & Dalton, J. (2009, July). Characterizing the subjectivity of topics. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (pp. 642-643).
41. Dewulf, A. (2013). Contrasting frames in policy debates on climate change adaptation. *Wiley Interdisciplinary Reviews: Climate Change*, **4**(4), 321-330.
42. Sentence-Transformers. (2022). *all-MiniLM-L6-v2*. Hugging Face. Retrieved from *huggingface.co*
43. Lockwood, M. (2018). Right-wing populism and the climate change agenda: exploring the linkages. *Environmental Politics*, **27**(4), 712-732.
44. Frischlich, L., Klapproth, J., & Brinkschulte, F. (2020). Between mainstream and alternative–Co-orientation in right-wing populist alternative news media. In Proceedings of the First Multidisciplinary International Symposium, MISDOOM 2019, Hamburg, Germany, Springer.