Linguistic Pattern Analysis in the Climate Change-Related Tweets from UK and Nigeria

Ifeoluwa Wuraola
University of Hull, School of Computer Science, Hull, UK
i.a.wuraola-2021@hull.ac.uk

Nina Dethlefs
University of Hull, School of Computer Science, Hull, UK
n.dethlefs@hull.ac.uk

Daniel Marciniak
University of Hull, School of Criminology, Sociology and Policing, Hull, UK
d.f.marciniak@hull.ac.uk

Abstract

To understand the global trends of human opinion on climate change in specific geographical areas, this research proposes a framework to analyse linguistic features and cultural differences in climate-related tweets. Our study combines transformer networks with linguistic feature analysis to address small dataset limitations and gain insights into cultural differences in tweets from the UK and Nigeria. Our study found that Nigerians use more leadership language and informal words in discussing climate change on Twitter compared to the UK, as these topics are treated as an issue of salience and urgency. In contrast, the UK's discourse about climate change on Twitter is characterised by using more formal, logical, and longer words per sentence compared to Nigeria. Also, we confirm the geographical identifiability of tweets through a classification task using DistilBERT, which achieves 83% of accuracy.

1 Introduction

The IPCC reported in 2022 that climate change is currently impacting all inhabited regions worldwide, with human activities contributing to many observed changes in physical and biological systems (IPCC, 2022). The electronic data produced by internet users during climate-related events can offer valuable insights into how different geographic areas perceive the risks associated with climate change (Vicari et al., 2019). However, intercultural dialogue and discourse are increasingly being studied in linguistics, as culture is seen as a fundamental aspect of human activity (Hong et al., 2003). One such area of study is conversation culturomics, which uses language analysis to understand human culture and can help conservationists respond to cultural trends while staying socially relevant. Previous research by Ladle et al. (2016) identified ways in which this analysis can be useful, such as assessing the cultural impact of conservation interventions and promoting public understanding.

Our research aims at contributing to the comprehension of the interactions that exist across the UK and Nigeria by analysing and identifying linguistic features that each group uses to communicate the climate change narrative and therefore gain insights into the factors that shape these opinions and identify areas where more education and conservation interventions are needed. This is relevant as previous research by Diehl et al. (2019) highlighted a focus on Anglophone culture in studying climate change and conceptions on social media. Therefore, this study's inclusion of African perspectives contributes to the overall understanding of cultural differences in climate-related discourse.

In this paper, we use linguistic feature analysis supported by transformer networks to enhance classification performance and generate insights on cultural differences in climate-related discourse. Seen as only roughly 2% of Twitter data is geo-tagged (Karami et al., 2021), our study is based on a small dataset, and we look to overcome the limitations of data sparsity when analysing specific cultures through the inclusion of linguistic and socio-cultural features.

Proceedings of the 2023 CLASP Conference on Learning with Small Data, pages 90–97
September 11–12, 2023.

2 Related Works

Researchers have contributed to the field of NLP by providing a wide range of approaches and techniques for analysing and predicting sentiments. Recent research by Tyagi et al. (2020) proposed a framework to examine the conversations on climate change between two communities on Twitter, namely activists and sceptics. The framework compares users' hashtags, bot percentage, and messaging to understand the differences between the communities. The study found that sceptics' messages focused more on attacking personalities, while activists' messages aimed to call for action against climate change. In addition to polarising tweet users, social media have been used to analyse emotions in tweets. Loureiro and Alló, (2020) studied climate policy opinions on Twitter in the UK and Spain. Results show UK sentiment is more positive, with anticipation prevailing, while fear is dominant in Spain. Gender analysis also indicates higher male tweeting in both countries, yet Spain demonstrates a more balanced gender distribution. Also, Hannak et al. (2012) research investigated sentiment patterns in tweets, particularly weather and time's impact on aggregate sentiment, and evaluated how clearly the well-known individual patterns translate into population-wide patterns. Machine learning techniques with weather-correlated tweets shows aggregate sentiment follows distinct climate, temporal, and seasonal patterns.

Furthermore, Chen et al. (2019) used deep neural networks to detect climate change skeptics from tweet content and analyzed Twitter's climate change discussions and influencing factors over time. They created a neural network model with an 88% accuracy in identifying deniers. Extreme weather events and policy shifts were noted to influence public interest and attitudes toward climate change.

Studies that systematically explore intercultural differences are rare. Liu and Zhao, (2017) show that NGOs in China typically work to frame climate change within national Chinese context, to highlight the relevance and impact for all Chinese people. This common perspective is typical in collectivist cultures (Diehl et al., 2021). In contrast, African NGO outlets often frame climate change discourse in terms of increased agricultural hardships and social hardships and place an emphasis on educating citizens on the matter (Ford et al., 2015). Education is also a topic in other Asian countries, including Afghanistan, Bhutan, Kiribati, Nepal, and Tuvalu, where NGOs additionally work towards implementing adaptive strategies to climate change risks and increasing scientific scholarship (McGregor et al., 2018). All these studies are based on social science methodology and often rely on extensive manual annotation. In this paper we want to explore how linguistic features can overcome the burden of annotation. Also, Schäfer and Painter, (2021) contrast climate journalism in the global north and south. The authors find that while climate coverage has changed globally over the last decade to move increasingly online, there are fewer journalists who specialise specifically on climate in the global south, which impacts public information in those regions accordingly.

This paper proposes a new method to analyse linguistic features and cultural differences in climate-related tweets. Our study combines transformer networks with linguistic feature analysis to address small dataset limitations and gain insights into cultural differences. We conclude by highlighting the importance of understanding cultural differences, particularly in the Nigerian/African perspective, in climate discourse to facilitate effective action against climate change. The focus on the Nigerian/African perspective adds a novel contribution to the existing literature.

3 Methodology

3.1 Data Collection and Curation

Key words	Nigeria	UK
“climate change”	*“my believe is that climate change will eventually lead to lives extinction”*	*“net zero is a hoax that nobody is falling for manmade climate change is a lie and is impossible”*
“global warming ”	*“climate change global warming is a scam tho”*	*“yes to solve global warming is not about carbon banks and not filling ones kettle society will have to take look at itself and decide what kind of future we our children and grandchildren will have”*

Table 1: Sample tweets across Nigeria and UK based on key words.

Twitter has provided a wealth of data for analysis with 1 billion monthly visitors and 313 millions active users (Li et al., 2013), including topics such as climate change. Therefore, the datasets for this research were gathered from the Twitter API between September 2010 to April 2023. Tweets were filtered based on the key words shown in Table 1.

A total of 81,507 tweets were collected for the analysis, comprising 44,071 from the UK and 37,436 from Nigeria. Also, only English-language tweets (since the English set is the focus of this project) were kept. These are identified by running the tweets through a langdetect algorithm (a model that identifies the language used in the text within the specified range).

3.2 LIWC

Our study uses Linguistic Inquiry and Word Count (LIWC-22) to perform linguistic analysis on geo-tagged data collected from Twitter. LIWC-22 is a software developed by Boyd et al. in 2022. It analyses word use within a text and calculates the percentage of word use for certain linguistic categories. A study in Indonesia had previously used LIWC to filter names from tribe, religion, and race and observed that the use of names is mostly followed by negative sentiments (Adi and Eka, 2021). This paper uses LIWC to investigate the linguistic differences in climate change discourse in English tweets from the UK and Nigeria. The LIWC analysis produced 124 variables of word categories, and the top 20 variables with the highest variance were selected for further analysis. In addition to linguistic features, valence (degree of positiveness or negativeness) was also calculated using VADER (Valence Aware Dictionary for sEntiment Reasoning), a sentiment analysis tool commonly used for analysing social media data (Hutto, 2022).

3.3 Classification Task

Our study proposes a transformer network to classify country distribution based on geo-tagged tweets to verify the geographical origin of the tweets. As fine-tuning BERT can be challenging due to its complex structure and parameters, we use DistilBERT created by Sanh et al. (2020), a compressed version with fewer parameters that is easier and faster to fine-tune with moderate resources. The study explores the use of DistilBERT in pre-training and fine-tuning to build the model.

4 Results and Discussion

4.1 Linguistics Feature Analysis across Nigeria and UK

For each tweet, we take the average of the valence and LIWC values of each of the words in the tweet

	Description	Nigeria	UK	p-value
Valence	Positiveness-negativeness	0.17	0.13	<0.0001
	LIWC variables			
Tone	Degree of positive (negative) tone	41.51	40.48	<0.0001
Authentic	Perceived honesty, genuineness	31.79	31.49	>0.0001
Clout	Language of leadership, status	58.14	31.490	<0.0001
Analytic thinking	Metric of logical, formal thinking	72.96	75.13	<0.0001
Linguistics	Linguistics dimensions	45.78	46.86	<0.0001
Pronoun	I, you, that, it	7.53	6.92	<0.0001
Determinants	the, at, that, my	9.05	9.51	<0.0001
WPS	Average words per sentence	28.88	31.46	<0.0001
WC	Total word count	28.88	31.46	<0.0001
Cognition	Psychological processes (is, was, but, are)	9.25	8.84	<0.0001
Social	Social processes	9.32	7.61	<0.0001

Big words	Percent words 7 letters or longer	27.61	28.10	<0.0001
Cogproc	Cognitive processes (but, not, if, or, know)	8.34	7.93	<0.0001
Perception	in, out, up, there	7.86	8.37	<0.0001
Verb	is, was, be, have	9.88	9.57	<0.0001
Dictionary words	Percent words captured by LIWC	64.69	64.76	>0.0001
Function	the, to, and, I	35.09	35.26	>0.0001
Adjective	more, very, other, new	5.34	6.37	<0.0001
Preposition	to, of, in, for	10.53	10.76	<0.0001
Allure	have, like, out, know	5.82	5.13	<0.0001

Table 2: Description of variables, average valence, LIWC scores, and p-value at country level.

text. The average of the variables was computed at the country level. Table 2 displays the valence and 20 selected variables for the study, with the highest value being highlighted, and presenting the distribution p-value for each. We observe the average valence of Nigerian tweets is higher (more positive) than the tweets from the UK.

4.2 Interpretation of Findings

According to the study, Nigerians tend to use more leadership language in climate change discourse on Twitter compared to the UK, which may be due to cultural and social norms. Nigerian culture places more emphasis on leadership and authority figures, given the country's political instability and history of leadership challenges in climate change and environmental degradation (Uyigue and Agho, 2007). As a result, climate-related tweets in Nigeria may be targeted toward government officials or policymakers, leading to the use of more leadership language to convey a sense of urgency and importance. Examples of tweets to support this claim are "*save niger delta environment from pollution extinction of sea foods activist pleads portharcourt the federal government and indeed the people of the niger delta have been urged to pull resources together amp save the region s environment from further pollution*" and "*climateaction requires significant investments by governments and businesses but climate inaction is vastly more expensive lets all unite to take climate actions now*". In addition to leadership, Nigeria's vulnerability to the impacts of climate change may be reflected in the use of emotional and informal language when discussing climate-related matters (Adelekan, 2010), making such issues more salient and urgent to them. Also, the study finds that Nigerian climate-related tweets have a higher frequency of verbs and pronouns compared to UK tweets. This may be because leadership is often expressed through specific linguistic features in text, such as the use of certain pronouns and verbs. (Tweets above use verbs such as 'save', 'take' which could be an attempt to convey a sense of urgency to the government). Furthermore, cognitive terms like "but", "if", etc, and social terms like "you", "we", "he", etc, are also higher in tweets from Nigeria compared to the UK.

The study finds that the discourse about climate change on Twitter in the UK is characterised by using 'bigger' and more formal/logical words, frequent use of adjectives and prepositions, and longer words per sentence compared to discourse from Nigeria. This difference could be attributed to the demographics of the authors of tweets, including NGOs and news outlets with more technical knowledge in climate science or related fields. This results in the use of specialised and technical lexicons, which is more evident in the UK than in Nigeria. Examples of tweets to support the claim are "*discovering different life perspectives could be just around the* *corner join climate solutions book club to make new friends from around the world and start exploring we read one book a month on climate change and then meet up on zoom to discuss https*" and *"human tiger conflicts seen to rise as migrants move into nepal national park conservationists raise concerns that the growing human presence in the chitwan district will pose additional challenges to conservation efforts https*" Additionally, the cultural emphasis on politeness and formality in the UK could also influence language use on social media (Sifianou, 1999). Another possible factor is that the urgency of climate change issues may not be as immediately felt or apparent in the UK as the nation is developed and her government is more proactive in addressing climate change policies compared to the Nigerian government, leading to a more detached and analytical discussion (O'Neill and Nicholson-Cole, 2009; Loureiro and Alló, 2020; Vu, 2020).

To further investigate individual differences in formal and informal language use, a cluster analysis was conducted on the dataset using the KMeans++ algorithms with 2 clusters, valence, and 20 LIWC variables as the experimental setup. The goal was to further explore formal-informal linguistic differences and observe whether clusters would automatically divide along this dimension. Specifically, we wanted to understand whether UK tweets were generally more formal than Nigeria tweets, or whether the type of user account that puts out a formal tweet is just more frequent in the UK than Nigeria. This is in line with research by Hopke and Hestres, (2018), who analyse the social media coverage during COP 21 (Paris) climate talks by different stakeholders. The authors show that while idiosyncrasies exist at a national level, pro-climate stakeholders, such as mainstream media outlets, NGOs, and prominent activists, showed notable similarities in the way they communicated about climate change and risks across countries and continents. Their analysis is manual and based on framing but seems in line with our linguistic findings. At a more detailed level, these findings imply that categorization of tweets using linguistic patterns and word use might assist environmental stakeholders in gaining knowledge of the target location to address scepticism about climate change and identify regions that require more education and advocacy.

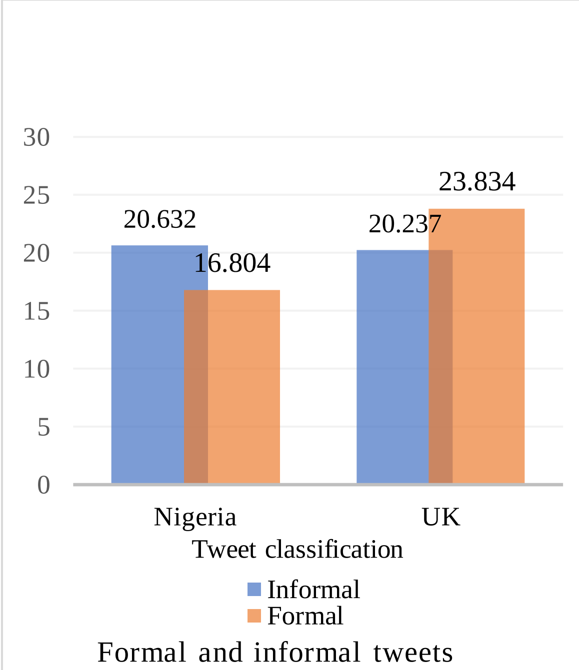

Figure 1: A figure showing Tweet classification across Nigeria and UK.

4.3 Experimental Results and Interpretation (classification tasks)

For our classification study, we randomly set the training size to 80% and we conduct experiments using 2 layers, 100 hidden units, dropout of 0.1, learning rate of 0.0001, Adam optimisation and 10 epochs. We achieved 83% of accuracy from the DistilBERT model when predicting the originating country of a tweet. We present qualitative analysis of a set of examples in Table 3. As can be seen, our classifier was able to correctly predict the country origin of 4 out of 5 non-geotagged tweets, indicating its potential use in identifying the origin of climate-related tweets. This can help policymakers in promptly addressing climate change disasters in specific regions. The model's predictions suggest that certain linguistic features may have played a role in predicting the country origin of tweets. As discussed in section 4.2, use of longer and bigger words per sentence is more common in the UK's climate-related discourse than in Nigeria, and this could have affected the model's ability to make accurate predictions. An example is the third non-geotagged tweet in Table 3, which had longer words per sentence and was wrongly predicted as originating from the UK. The accuracy of the study's predictions may have been influenced by factors such as dataset size, bias in training data,

Non-geotagged tweet	Actual label	Predicted label
'i m in love with nature'	Nigeria	Nigeria
'no you said there is no link and deforestation is solely a problem for overpopulation'	UK	UK
'revealed rampant deforestation of amazon driven by global greed for meat https t co'	UK	UK
'it is a very good season with a lot of rain so please plant any native flora it will be a great service to nature here you can see the benefits of trees biodiversity native'	Nigeria	UK
'what a nice gift from nature why bother it greeniewo savetheplanet gogreenie'	UK	UK

Table 3: Classification performance of non-geotagged tweet.

or other issues. A dataset including tweets from a wider variety of countries and regions could help to better train the model to recognize linguistic patterns that are specific to different cultural contexts. Also, testing the model on different types of tweets, such as those related to other environmental issues or different types of disasters, could help to determine whether the model's performance is specific to climate-related tweets or whether it can be applied more broadly. Therefore, further research could be done in those areas to optimise the model's accuracy.

5 Conclusion

The study employed linguistic feature analysis supported by transformer networks to investigate cultural differences in climate-related discourse between Nigerian and British tweets, aiming to identify trends in their respective lexicons. Through this approach, we were able to achieve a respective baseline classification performance and mitigate the limitations of working with small datasets. Our findings suggest that studying linguistic patterns and word use are crucial areas of research in socio-cultural analysis tasks, particularly in the classification of tweets based on their location. This is particularly useful in quickly addressing climate change disasters in specific geographic areas, as well as gauging public interest in climate change, and characterising discourse in different cultures even with limited data availability. Our experiments with DistilBERT on small data yielded promising results, with an accuracy of 83% in correctly classifying the country of origin for climate discourse tweets.

Based on these findings, we could recommend that those interested in identifying the country of origin of climate discourse tweets using linguistics patterns should focus on language that conveys a clear positive or negative sentiment and complex language in a way that is perceived as authoritative. Additionally, while the level of analytical thinking and linguistic complexity may also be important in predicting the country of origin of climate discourse tweets, they may not have as significant an impact as the overall sentiment and complexity of language. Future research will examine tweets in languages other than English as well as tweets from other countries. In addition to the linguistics pattern, we will look at the emotional dynamics surrounding climate change in these countries over an extended period.

Limitations

Although the present research identified patterns in the linguistic features of tweets, it only analysed English tweets. Therefore, the multilingualism of Twitter users should be considered to gain deeper insights into linguistic patterns and word use, as this can improve the analysis and prediction of such patterns. Also, given the scope of this research, it's important to acknowledge its limitation in definitively establishing whether the observed cultural disparities are attributed to climate change or general language differences. This opens the door for future investigations, potentially applying our methods to other disciplines for broader insights. Furthermore, given the potential limitation that geotagging may lead to a non-representative Twitter sample, our research's utilization of DistilBERT for training becomes pivotal in addressing this concern. Thus, the integration of non-geotagged tweets into the classifier presents a promising direction for future investigations, effectively addressing potential limitations.

Ethics Statement

The study followed the ACL Ethics Policy to ensure ethical and responsible conduct throughout the research process. We collected and analysed publicly available tweets, ensuring privacy and confidentiality of the Twitter users. Also, we avoid perpetuating stereotypes or biases and conducted the research in a respectful manner that aligned with cultural norms and values. Informed consent was not required since the data was publicly available, but we anonymized the data to protect individual users. The study uses appropriate statistical and computational methods and shared our findings transparently with the wider research community. We are committed to upholding ethical principles in their research.

Acknowledgments

The authors express gratitude to the Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM) for generously funding and enabling this research.

References

Ibidun O Adelekan. 2010. Vulnerability of poor urban coastal communities to flooding in Lagos, Nigeria. *Environment and Urbanization*, 22(2):433–450. https://doi.org/10.1177/0956247810380141.

Ryan Boyd, Ashwini Ashokkumar, Sarah Seraj, and James Pennebaker. 2022. *The Development and Psychometric Properties of LIWC-22*. https://doi.org/10.13140/RG.2.2.23890.43205.

Xingyu Chen, Lei Zou, and Bo Zhao. 2019. Detecting Climate Change Deniers on Twitter Using a Deep Neural Network. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, pages 204–210, New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3318299.3318382.

Trevor Diehl, Brigitte Huber, Homero Gil de Zúñiga, and James Liu. 2019. Social Media and Beliefs about Climate Change: A Cross-National Analysis of News Use, Political Ideology, and Trust in Science. *International Journal of Public Opinion Research*, 33. https://doi.org/10.1093/ijpor/edz040.

Trevor Diehl, Brigitte Huber, Homero Gil de Zúñiga, and James Liu. 2021. Social Media and Beliefs about Climate Change: A Cross-National Analysis of News Use, Political Ideology, and Trust in Science. *International Journal of Public Opinion Research*, 33(2):197–213. https://doi.org/10.1093/ijpor/edz040.

Brett Q. Ford, Julia O. Dmitrieva, Daniel Dmitrieva, Yulia Chentsova-Dutton, Igor Grossmann, Maya Tamir, Yukiko Uchida, Birgit Koopmann-Holm, Victoria A. Floerke, Meike Uhrig, Tatiana Bokhan, and Iris B. Mauss. 2015. Culture shapes whether the pursuit of happiness predicts higher or lower well-being. *Journal of Experimental Psychology: General*, 144(6):1053–1062. https://doi.org/10.1037/xge0000108.

Aniko Hannak, Eric Anderson, Lisa Feldman Barrett, Sune Lehmann, Alan Mislove, and Mirek Riedewald. 2012. Tweetin' in the Rain: Exploring Societal-Scale Effects of Weather on Mood. *Proceedings of the International AAAI Conference on Web and Social Media*, 6(1):479–482. https://doi.org/10.1609/icwsm.v6i1.14322.

Ying-Yi Hong, Veronica Benet-Martinez, Chi-Yue Chiu, and Michael W. Morris. 2003. Boundaries of Cultural Influence: Construct Activation as a Mechanism for Cultural Differences in Social Perception. *Journal of Cross-Cultural Psychology*, 34(4):453–464. https://doi.org/10.1177/0022022103034004005.

Jill Hopke and Luis Hestres. 2018. Visualizing the Paris Climate Talks on Twitter: Media and Climate Stakeholder Visual Social Media During COP21. *Social Media + Society*, 4:205630511878268. https://doi.org/10.1177/2056305118782687.

C. J. Hutto. 2022. cjhutto/vaderSentiment.

IPCC. 2022. The Sixth Assessment Report of the IPCC | UNFCCC.

Amir Karami, Rachana Redd Kadari, Lekha Panati, Siva Prasad Nooli, Harshini Bheemreddy, and Parisa Bozorgi. 2021. Analysis of Geotagging Behavior: Do Geotagged Users Represent the Twitter Population? *ISPRS International Journal of Geo-Information*, 10(6):373. https://doi.org/10.3390/ijgi10060373.

Richard J Ladle, Ricardo A Correia, Yuno Do, Gea-Jae Joo, Ana CM Malhado, Raphaël Proulx, Jean-Michel Roberge, and Paul Jepson. 2016. Conservation culturomics. *Frontiers in Ecology and the Environment*, 14(5):269–275. https://doi.org/10.1002/fee.1260.

Linna Li, M. Goodchild, and Bo Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. https://doi.org/10.1080/15230406.2013.777139.

John Liu and Bo Zhao. 2017. Who speaks for climate change in China? Evidence from Weibo. *Climatic Change*, 140. https://doi.org/10.1007/s10584-016-1883-y.

Maria L. Loureiro and Maria Alló. 2020. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. *Energy Policy*, 143(C).

Ian McGregor, Hilary Yerbury, and Ahmed Shahid. 2018. The Voices of Local NGOs in Climate Change Issues: Examples from Climate Vulnerable Nations. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 10:63–80. https://doi.org/10.5130/ccs.v10.i3.6019.

Saffron O'Neill and Sophie Nicholson-Cole. 2009. "Fear Won't Do It": Promoting Positive Engagement With Climate Change Through Visual and Iconic Representations. *Science Communication*, 30(3):355–379. https://doi.org/10.1177/1075547008329201.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108 [cs]. https://doi.org/10.48550/arXiv.1910.01108.

Mike Schäfer and James Painter. 2021. Climate journalism in a changing media ecosystem: Assessing the production of climate change-related news around the world. *Wiley Interdisciplinary Reviews: Climate Change*, 12. https://doi.org/10.1002/wcc.675.

Maria Sifianou. 1999. *Politeness Phenomena in England and Greece: A Cross-Cultural Perspective*. Oxford University Press. Google-Books-ID: hy1RCbiyWYkC.

Aman Tyagi, Matthew Babcock, Kathleen M. Carley, and Douglas C. Sicker. 2020. Polarizing Tweets on Climate Change. arXiv:2008.13054 [cs].

Etiosa Uyigue and Matthew Agho. 2007. Coping with Climate Change and Environmental Degradation in the Niger Delta of Southern Nigeria.

Rosa Vicari, Ioulia Tchiguirinskaia, Bruno Tisserand, and Daniel Schertzer. 2019. Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience. *Natural Hazards and Earth System Sciences*, 19(7):1485–1498. https://doi.org/10.5194/nhess-19-1485-2019.

Hong Vu. 2020. Social media and environmental activism: Framing climate change on Facebook by global NGOs. *Science Communication*, Accepted.