

Climate change and 'climategate' in online reader comments: A mixed methods study

Introduction

In parallel with print 'letters to the editor' as an outlet for reader views, newspapers currently provide a range of opportunities for audiences to communicate their opinions and discuss issues in real time. Although research shows that such Internet-based opportunities remain under-utilised especially by tabloid readers (Richardson and Stanyer 2011), the issue of climate change has recently attracted a relatively large number of comments across UK newspaper platforms. For scholars of the sociology of science, science communication, and media studies such texts represent a unique window into spontaneous and creative non-expert conceptualisations of climate change by Internet users. Moreover, such discussions are arguably one of the best sources to study such dimensions of the climate change issue as ethics, morality and uncertainty, as online debates show how different actors attempt to redefine existing constructions of various problems and solutions (Koteyko 2010). In this study, we focus on one type of such web-based data that so far has not been explored in relation to climate change – online reader comments. We will present results of our analysis of comments published on the website of the Daily Mail, the UK's second biggest-selling newspaper.

Online publishing often means easy access and few restrictions on the size of online contributions translating into the vast volume of texts that present methodological challenges in terms of data analysis. In this paper we therefore also want to discuss a mixed methods approach to the analysis of such web-based data. The proposed approach draws on corpus-assisted discourse analysis (Partington 2010; Baker et al 2008) that employs a combination of the quantitative study of word frequencies and qualitative examination of context, and is complemented by statistics mined from the source website. The qualitative component is based on the adaptation of the discourse analytic framework developed by Reisigl and Wodak (2009) to the constraints of web-based interactions (Herring 2004). The approach can be applied to provide timely insights into the framings of climate change in different online platforms following major mediated events, e.g. political milestones such as international agreements or indeed controversies such as 'climategate' and a plethora of other 'gates' that have repeatedly appeared on the climate change scene.

Background and research aims

In November 2009, documents and digitally stored email correspondence between the Climatic Research Unit leader Professor Phil Jones and his fellow researchers at the University of East Anglia were published on the Internet, just weeks before the United Nations Copenhagen Summit. The data and emails were obtained from a back-up server at the Climatic Research Unit following Freedom of Information requests in July the same year (Holliman 2011). The content of these emails was then used as a basis to argue that scientists were manipulating and withholding the data that disproved the severity of climate change, leading to the controversy that was dubbed 'climategate' in the media and blogosphere (Hoffman 2011). Subsequent inquiries cleared the scientists of

wrongdoing, but at the same time called for more transparency regarding data collection and analysis.

Since then a number of scholars have grappled with what this episode means for climate change communication (Hoffman 2011; Nerlich 2010), whereas climate scientists themselves speculated about its legacy, identifying major shifts in 'how climate science is conducted, how the climate debate is framed and how climate policy is being formed' (Hulme 2010). Hoffman (2011, 7), for example, notes that 'prominent climate deniers' quickly capitalised on the controversy to challenge the notion that climate change is a legitimate problem. At the same time, media scholars have extensively examined how various newspapers, and tabloids in particular (Boykoff 2008), deal with the issue of climate change; whereas opinion polls registered shifts in public perceptions of climate change both in the US (Leiserowitz et al in press) and the UK (Reser et al 2010).

However, so far no study has explored how newspaper readers have engaged with this issue, and how any future climate change-related topics can be studied by science communication scholars in this online space. Holliman (2011, 832) comes close to this question by speculating that the East Anglia controversy may have 'profound implications' for how climate science is discussed in the digitally mediated public sphere, but does not study actual online contributions. In contrast, this study aims to provide a combined quantitative and qualitative analysis of climate change-related topics discussed by tabloid readers in the year following 'climategate'. To reveal potential shifts in the way climate science is discursively framed by this particular sample of readers we will compare the 2010 online comments with those published on the Daily Mail website in the year immediately preceding the controversy. We suggest that if the hacking of Climatic Research Unit emails, the distribution through online socio-technical networks, and its media framing as 'climategate' has left an impact on readers' conceptualisations of and views on climate change, this will manifest itself both through qualitative changes (choice of lexis, use of different words) and through quantitative shifts (frequencies of word use).

We focus on the following research questions:

1. What are the frequent topics of, or issues discussed in, comments to tabloid articles dealing with climate change in 2010?
2. How do these online contributions compare to reader comments on climate change written before November 2009 (in terms of content and main themes)?
3. How is climate science discursively constructed by the tabloid readers following the 'climategate' controversy?

The secondary, and methodologically oriented, aim is to demonstrate how techniques of corpus-assisted analysis can be employed to interrogate increasingly large data sets generated online on the topic of climate change, and in this way complement more qualitative social science frameworks. Although the analysis is based, in part, on word frequency and word distribution provided by a specialist software (Scott 2011), we are not concerned with opinion summarisation (e.g. Potthast and Becker 2010), but aim to provide an empirically grounded approach combining quantitative and qualitative techniques.

Data

The website archive of The Daily Mail, the UK's second biggest-selling tabloid was searched for the term 'climate change' (www.thedailymail.co.uk). We then strategically sampled all available comments to articles published between January and December 2010 (Corpus 1: 4698 comments, 414, 462 words,) and between November 2008 and November 2009 (Corpus 2: 1799 comments, 148, 895 words). It should be noted that not all articles returned through the Daily Mail search facility as containing the phrase 'climate change' attracted comments, or were open for comments. For example, out of 355 news articles on climate change listed on the website, only 233 appear to have generated climate change-related posts[1]. We also do not know how many, and what kind of, comments were removed as a result of moderation.

The comments are not representative of public opinion on climate change, but reflect the views of those with access to the Internet who tend to read online news. There is also a possibility that there is a disproportionate participation by readers who have a strong interest in the topic of climate change.

Conceptual framework

In recent years, social sciences have undergone a 'discursive turn' marked by an increased interest in the role of language played in the creation of our social reality (Bhatia et al 2008 eds). Social studies of environment and climate change are not an exception, as according to Killingsworth and Palmer, "As much as the environmental dilemma is a problem of ethics and epistemology, it is also a problem of discourse" (Killingsworth and Palmer 1992, 6). In contrast to the abundance of studies dedicated to the analysis of print media discourses (Boykoff 2008; Carvalho 2007; Koteyko 2012), social science scholars have so far paid little attention to constructions of climate change in online spaces, and online reader comments in particular. Like media reporting on climate change, reader comments on this reporting contribute to the diverse, complex and contested social representational field, in which climate change is embedded, and thereby have the potential to shape thinking and public discourse around this complex environmental phenomenon (Jaspal et al under review). As such, online comments can be potentially approached from the perspective of Social Representations Theory (Moscovici 1988) or Frame theory (Benford and Snow 2000) which share the aim of gaining an understanding of how language functions in constituting and transmitting knowledge, although the large volume of online data remains a major challenge. In this article we therefore seek to demonstrate how a corpus-assisted approach developed for analysis of large archives of electronically stored texts or corpora (Stubbs 2001; Partington 2010) can provide a grounded study of online reader data, and at the same time pinpoint areas of interest for a subsequent detailed analysis within these frameworks.

At the same time, those sociological studies that do pay attention to Internet-based discussions on climate change (Lederbogen and Trebbe 2003) still tend to rely on a type of content analysis that recruits a set of concepts defined by the analyst rather than participants themselves, and also involves classifying statements for what they are rather than what they do (Lamerichs and te Molder 2003). Such an analysis inevitably misses out on local interactional business that participants may attend to in online spaces, such

as defending, undermining, constructing and maintaining authority and so on, and rarely takes into account the specific affordances of the medium (Hutchby 2001). To overcome this limitation the approach to online reader comments adopted in this article is therefore informed by a linguistic perspective and is closest to what Herring (2004) refers to as 'computer-mediated discourse analysis'. Our framework draws on methodological paradigms that originated in the study of spoken and written communication, such as corpus linguistics, pragmatics, and critical discourse analysis but also allows the inclusion of contextual parameters specifically related to computer-mediated communication, such as lack of audio-visual cues and structural affordances of the source website. When applied to online reader comments, which in a previous study (Laslo et al 2011) were found to refer two to three times more to preceding comments rather than to a news article on a scientific issue, such analysis can help reveal how 'top-down' print media frames and discourses may be reproduced or resisted locally in lay talk, and within the specific parameters of a web-based interactional context.

Previously, corpus-based techniques were predominantly employed in collections of text counting millions of words in order to make generalisations about meanings and grammatical constructions in language as whole. More recently, however, the approach of corpus-assisted discourse studies espoused by Partington (2003; 2010) drew attention to the fact that a study of lexical patterns and how they differ in selected, often much smaller and purpose-built corpora can also help the researcher examine changes in 'discourse processes' that can be conditioned by 'external social influences' (2010, 86). As we hope to demonstrate below, the study of discourses with corpus linguistic techniques can reveal the politics of naming and framing in the struggle over meaning, rather than semantic changes typically examined in relation to internal linguistic factors (Koteyko, 2007). The analysis does not stop at mere description of linguistic phenomena but involves discussion of wider issues relating to social practice. As such, it can be a useful tool in sociological studies seeking to examine how climate change is discursively constructed using language as a means to negotiate, support or challenge certain definitions and underlying assumptions (Alexander 2008).

Two analytical tools are central in the quantitative part of such analysis: *keywords* and *collocation*. Keywords are used to compare the *relative* frequency of words in any corpus with reference to another corpus. The first step is to draw lists of words in order of their absolute frequency using the 'Word list' function in WordSmith (Scott 2011), one for each corpus. The Keywords tool of the same software is then used to compare the contents of the two lists, relying on chi-squared or log-likelihood tests. Those words that are unusually frequent in the first list/corpus compared to the frequency of the same word type in the second list (or reference corpus) are called keywords and are themselves displayed as a list. The more statistically significant an item is, the more *key* it is (its keyness value is higher), and the higher it is placed on this new list (Partington 2010). The resulting keyword list therefore provides the analyst with words that are *salient* in one corpus compared to another corpus; it can be a useful tool for identifying lexical items which warrant further investigation (Baker 2006, 125). The procedure can be repeated by inverting the two corpora to reveal the items which are salient in the second corpus. *Collocation* is a lexical relation between two or more words which have a

tendency to co-occur within a few words of each other in running text (Stubbs 2001, 24). Analyzing collocation lists can reveal the semantic fields in which the word is used (collocational networks) and give a first glimpse of the rhetorical strategies coded implicitly and explicitly in language use. Such collocational networks can be in turn the first step into the meanings of social processes and descriptions of social actors constructed by a given group of language users (and which can be explored further following the abovementioned qualitative frameworks).

In addition to the above techniques, a qualitative corpus linguistic tool in the form of concordances will be also employed. Concordance analysis (see figure 3 for example) allows examination of the lexical environment of a search term (or node). As Baker et al (2008) observe, in a corpus-assisted study the researcher is normally required to analyse numerous concordance lines by hand to identify wider themes or patterns that may not be easily evident via collocation, keyword or frequency analysis. Overall then, keywords can provide comparative analysis as well as serve as a useful guide to popular topics, whereas collocates and concordances provide further contextualisation. A combination of both provides a 'feel' of the corpus and a 'map' of key junctions in texts under study (Baker et al 2008, 284). Last but not least, it is important to point out that as in any study using automation, no matter how good the computational tools are the challenge of data interpretation is always faced by the researcher. In a corpus-assisted study of discourse, this translates into the role of the analyst to make sense of the linguistic patterns generated by the software.

The qualitative analysis of concordances is informed by the approach developed by Reisigl and Wodak (2001) for the analysis of positive self-presentation and negative other-presentation, which these authors later applied to examine discourses on climate change and global warming (Reisigl and Wodak 2009). We also draw on linguistic studies of online discussions (Myers 2010; Herring 2004; Armstrong et al in press). Out of the vast arsenal of analytical tools offered by these authors, we have to limit the analysis to actors' description and stance-taking evident in strategies employed for predication, labelling, and intensification or mitigation (see Table 3 below). Each of these strategies is manifested textually through different linguistic (lexical) indicators, such as specific words used to achieve construction of in-groups and out-groups. Focusing on these categories will allow for a better comparison of the online readers' comments with the previous study of climate sceptic discourses (Reisigl and Wodak 2009).

Analysis

Keywords: mapping the points of interest

In line with research question 1, we carried out a keyword analysis to establish the main topics characterising corpus 1. The keyword analysis is commonly employed to provide an indication of *aboutness* (Partington, 2010). In our study, the overall topic is already known as only comments to newspaper articles on climate change were collected. We therefore employed keywords and concordances to learn more about actors and processes covered in comments, and the stance towards them. The examination included: examining single keywords and links between them (e.g. semantically grouping them in relation to specific topic); examining collocates of keywords; generating and

examining concordances for keywords or collocates.

As can be seen from Table 1, SCIENCE is the most key item on the 2010 list, closely followed by IPCC, DATA, AGW, JONES, BBC, PAPERS, SCIENTISTS, POST, CLIMATEGATE, and REPORTS. These keywords leave little doubt about the central topics of comments in 2010, providing an immediate, if somewhat crude, answer to research question 1. Grouping the keywords SCIENCE and SCIENTISTS we can reasonably assume that the significant bulk of online contributions is focused on climate science, whereas the high keyness value of AGW (anthropogenic global warming), JONES (in reference to the research unit leader Professor Phil Jones), and of course, CLIMATEGATE indicate that conceptualisations of climate science are likely to be closely linked to the East Anglia Controversy. Moving down the list, we can identify a group of keywords which warrant specific attention such as FUNDED, BIASED, CONSPIRACY, DENIER, FRAUD, VESTED, DISCREDITED.

Corpus 1	SCIENCE, IPCC, DATA, AGW, JONES, BBC, PAPERS, POST SCIENTISTS, CLIMATEGATE, REPORTS, GLACIERS, SHORT BIN, MET, SOURCES, EMAILS, INTEREST, CAMERON, FUNDED, PROFESSOR, RAIN, TRADING, TEXAS, CONSENSUS, RESULTS, BIASED, TURBINES, FIGURES FAMILY, QUOTE, UPDATED, KNOWLEDGE, CONSPIRACY DENIER, NATIONAL, VAPOUR, RECENT, ADMIT, DEAL FRAUD, PERSONAL, DEGREE, CAMBRIDGE, WARMEST OFFICE, EXPERT, PREDICTIONS, WATER, LIST, BUILT, GROUP, VESTED, NUMBERS, EXPOSED, FINANCIAL, STATES, BILLS, ANTHROPOGENIC, CONCLUSIONS, DISCREDITED
Corpus 2	TAX, MEAT, MILE, PAY, BROWN, ROAD, STERN, TRANSPORT, VEGETARIAN, CAR, CARS, SAVE, LORD EATING, PLANET, ROADS, FUEL, ANIMALS, DRIVE, COWS METHANE, DIET, PEOPLE, LIVE, POLICE, MOTORISTS, CONGESTION, MOTORIST, DRIVING, EAT, EMISSIONS, PAYING, PETROL, AMERICANS, GOVERNMENT, MILES

Table 1: Keywords (in order of decreasing salience)

Turning to our research question 2, we then examined differences between reader comments on climate change before (corpus 2) and after (corpus 1) the hacking of emails at the University of East Anglia by comparing keyword lists drawn for each corpus. As we can see from Table 1 displaying the words salient in the two corpora, in contrast to corpus 1 where comments focused predominantly on (climate) science and technology, comments in corpus 2 centred on issues around tax and topics to do with food and transport.

After examining the collocational list and concordances generated for TAX, it became evident that the issue of taxation is used as a basis to argue against climate change (e.g. *tax ruse*, *tax scheme*, *tax scam*, *tax con* see Figure 1) – the framing which, as discussed below, is also frequently used in comments from Corpus 1. In this corpus however, the actors in

this 'tax scam' are almost exclusively politicians or 'government' in general, whereas comments from Corpus 1 portray scientists as the main 'perpetrators'.

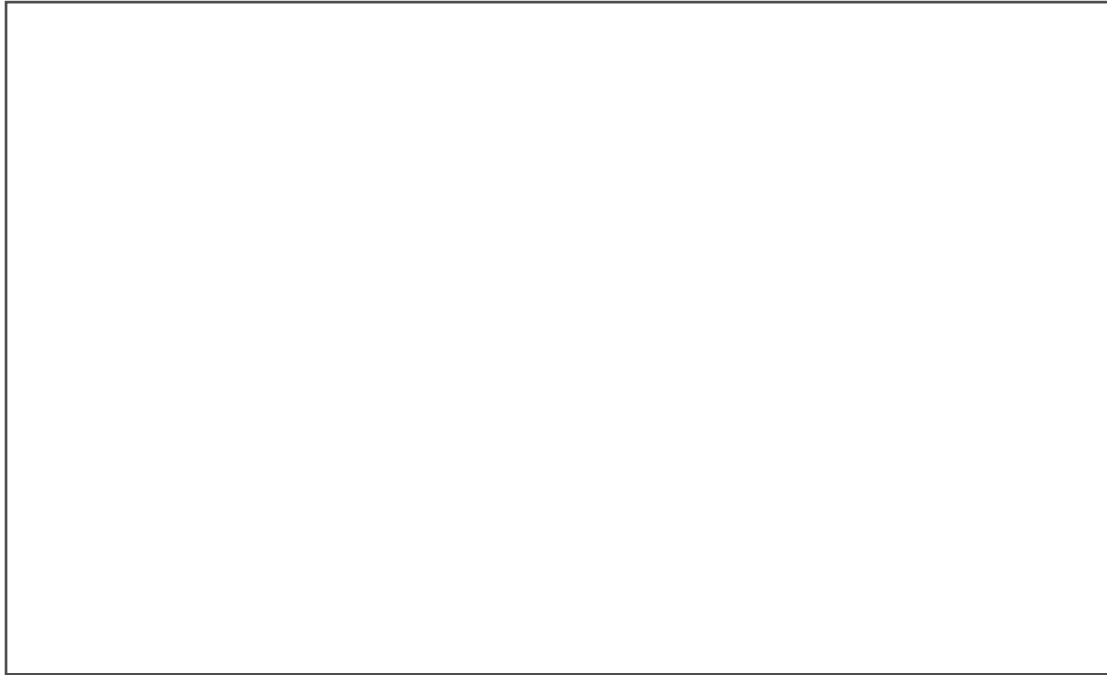


Figure 1: concordances of 'tax'

Collocates of 'science' and 'scientists: zooming in on specific arguments and their popularity

The high keyness value of *science* and *scientists* justifies the creation of a subcorpus to explore comments using these words in more detail. This also serves as data reduction strategy, as we can now concentrate on a limited amount of text (reduced from 4698 comments to 1907 concordance lines), knowing that it will contain the most salient lexical items out of the whole corpus. In fact, in a separate paper we examine these concordance lines and full comments within the framework of social representations theory (Jaspal et al under review). At the same time, we should be aware that this strategy inevitably leaves out some reader reactions, for example comments on global warming in general (e.g. 'Global warming a myth? just the same as low fares with Ryanair'). However, the analyst still has the means to capture such discourses by running additional concordances for the terms of interest.

Following research question 3, the corpus-based technique of collocation lists can bring us closer to the discursive means by which particular presentations of science and scientists are constructed in reader comments. As can be seen from Table 2, the adverb 'not', commonly used to express to negation, refusal or prohibition, is the top collocate and was used together with 'science' and 'scientist/s' 115 times and 75 times respectively. Although our analysis is focused on content words rather than grammatical words (prepositions, adverbs, articles), the consistently high co-occurrence of this adverb with several keywords (see also 'agw' in the next section) attracted our attention. As

Figure 2 shows, the use of ‘not’ signals a definitional battle of some sort, and an attempt to frame or re-frame existing conceptualisations of what constitutes a problem (it can be therefore examined further as an indicator of a ‘diagnostic’ frame, used to assess problems rather than solutions, Benford and Snow 2000).

By grouping together collocates relating to SCIENCE and SCIENTISTS (listed separately in Table 2), it became possible to build up a general picture of how climate science is constructed in corpus 1. For example, such collocates as *peer review*, *consensus*, *evidence* and *theory* alerted us to the possibility that readers may be drawing on what can be called a pervasive and established representation of science, characterised by rigour, objectivity, falsification and scepticism (Jaspal et al under review). One of the top collocates of scientists – *called* (29) – turned out to be part of a frequent word combination ‘so-called scientists’, the use of which points to a strategy of dissociating climate scientists from the ideal norm of science. Other collocates, such as *junk*, *pseudo*, *funding*, *paid* etc. point to a negative presentation, as something linked to fraud. Further on, such collocates as *politics*, *politicians*, and *government* echo the negative representations of climate change mitigation policies as a ‘tax scam’ in corpus 2.

Search term	Collocates (frequency)
SCIENCE*	not (115), settled (31), agw (29), real (28), understand (24), pseudo (23), review (15), theory (14), bad (13), called (13), man (13), peer (12), consensus (12), data (12), rocket (12), bbc (12), evidence (12), ipcc (11), money (11), reviewed (11), years (10), mean (9), point (9), politics (9)
SCIENTIST*	not (75), these (60), government (34), agw (30), called (29), work (23), data (21), politicians (21), IPCC (18), real (18), climategate (15), email (15), claim (14), grant (12), majority (11), qualified (11), scientific (11), world (11), Charles (10), funding (10), paid (10), sceptic (10), funding (10)

Table 2: Collocates of ‘science’ and ‘scientists’

The analysis of collocational lists enables a quick way of establishing frequent lexical patterns that may be indicative of broader themes and arguments. It provides assurance that discursive constructions emerging from the analysis are repeatedly used in the sample, and in this way helps to address a criticism often voiced against qualitative approaches: ‘The hidden danger is that the reason why the texts concerned are singled out for analysis in the first place is that they are not typical, but in fact quite unusual instances which have aroused the analyst’s attention’ (Koller and Mautner 2004, 218). However, to investigate what kind of (climate) science and scientists these collocational

networks are actually recruited to construct in our data, we need to turn to concordances and closer reading.

Concordances and the full text view: revealing discursive strategies

A systematic study of both standard concordances, e.g. Figures 1-4, and extended concordances (which amount to several lines of text and often reveal most of the comment under study) enabled an in-depth analysis of how particular descriptions of social actors are achieved. The WordSmith software (Scott 2011) allows sorting the lines in various ways – in alphabetic order, by the word to the right or left of the search term – or running concordances in the context of a word of interest (e.g. to see how the keyword ‘scientist’ is actually used with the top collocate ‘called’ (Figure 4). The concordances of ‘science’, for example, make immediately visible a variety of devices that can be explored further as traces of different discourses and arguments: inverted commas and the use of ‘so-called’ to signal distancing from a proposition, the emphatic use of NOT written in capitals, ‘Science’ with capital S, and with the definite article, and so on.

[pic]

Figure 2: Concordances of ‘science’

Concordances can also help us pinpoint the lexicalisation of a particular association (in this case, negativity) that may not be picked up through collocational lists, or that may be only hinted at by the proximity of less frequent lexis. For example, the top collocate of ‘science’ and ‘scientist/s’ - the acronym AGW - appears to be a neutral label given its origin in scientific literature, and its top collocates are *not* (55), *theory* (33) and *evidence* (17). However, the existence of such collocates as *scam* (15) and *believers* (14) together with the word combinations such as *agw crowd*, *agw camp*, *agw agenda*, *agw lobby*, *agw movement*, *agw mythologists*, *agw hypocrisy*, and *agw nuts* revealed through the study of concordances shows that in this particular corpus it is used with predominantly negative associations. At the same time, extended concordances can help us study how some readers contest and challenge this negative evaluation of climate scientists; such instances of contestation are unfortunately infrequent in this collection of comments and therefore unidentifiable through collocates or keywords (but see example [5] below).

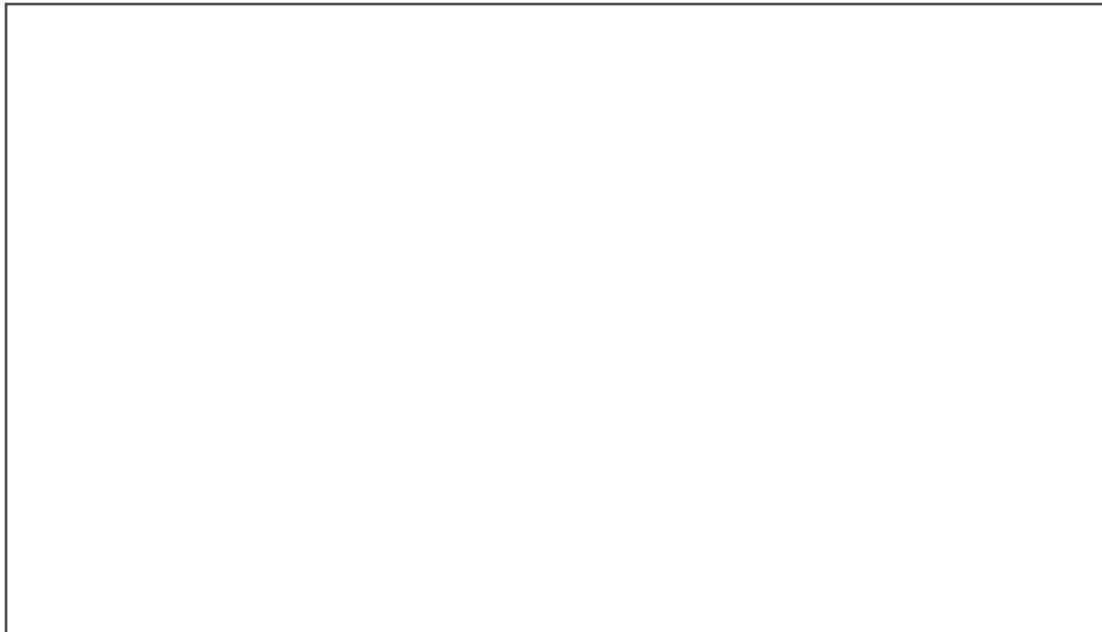


Figure 3: concordances of 'scien*' in the context of 'not'

Lexical patterns emerging from collocational lists and concordances were organised via discourse analytical categories of rhetorical strategies of self and other-presentation (Reisigl and Wodak 2009; Baker et al 2008), limited in this paper to the categories of actors' descriptions and stance. As a detailed discourse analysis of comments is beyond the scope of this paper, in Table 3 we present only a summary of findings. The strategies emerged from a process of self-presenting the views of the in-group (in this case people not convinced by science that climate change is occurring or not convinced of attributing the cause of climate change to human behaviour) by portraying the out-group as inept, corrupt and manipulative (for more detail, see Jaspal et al under review). Table 3 below details how the strategies of actors' description rely on two mechanisms: disassociation of climate scientists from category 'science' through the use of negation and definitions of 'real' scientists on the one hand, and attribution of negative traits to climate scientists through associations with politics or fraud/corruption (or both) on the other.

Strategies	Aim	Devices	Examples from Corpus 1
Referential/ Nomination	Construction of in-groups and out-groups	-membership categorization	(not) 'real scientists', 'bona fide scientists', pseudo-scientists, AGW nuts, AGW scam, the AGW propaganda, etc.
	In our data: to disassociate climate scientists from category 'science'	-labelling -references to scientific categories (see collocates)	'you don't have to be AGW believer to be wary of anecdotal evidence' 'global warming is theory, unrepeatable in a lab, yet taken as a scientific fact...'
Predication	Labeling social actors more or less positively or negatively	-stereotypical, evaluative attribution of negative or positive traits	'these money grabbing grant taking scientists.'; 'globalwarming scaremongering

	In our data: to attribute negative traits to climate scientists	(see concordances)	scientists and politicians'
Intensification	Modifying the epistemic status of a proposition	-intensifying or mitigating the illocutionary force or (discriminatory) utterances	'Consensus' IS NOT Science ! Never forget that.' ...are really not scientists
or mitigation			...do these supposed scientists really think that.. '...so-called AGW 'experts' 'so-called scientists'

Table 3. Strategies of negative other-presentation (adapted from Reisigl and Wodak 2009, 112-113)

One delegitimation strategy is to support the main tenets of science and then use this to argue that current climate science which provides evidence for anthropogenic global warming, or AGW, is not of this ideal type. Such a discursive construction of two types of science allows the commentator to proceed to a far from neutral contrast and comparison, achieved through statements differentiating between 'real' or even 'proper' science and 'agw science', used synonymously with 'pseudo'- and 'junk-science'. Here climate science is described as lacking the well-known attributes of scientific enquiry such as replicability, objectivity, or falsifiability and therefore labelled as *fraud, fiction, pseudo-science, conspiracy, swindle, scam* etc.

[1] This is a very **dodgy science** and flies in the face of **established scientific practice**. They just got swept into this ridiculous carbon footprint, global warming, CO2 hype.

[2] Agw science is not only not **settled**, it is pure fiction

As we can see from Table 3, another strategy proceeds by mapping dishonesty and greed typically associated in tabloid discourse with politics (especially politics around carbon credits and carbon tax) onto climate science. The following comment is a good illustration as it combines both strategies:

[3] It is standard practice in every **proper science** to release date and methodology in the greatest of detail so that every aspect of the research and of the argument can be 'falsified' (using the Popper meaning of the word). [...]. Not to do so puts climate research at the level of iridology, homeopathy, and alchemy. Add political agenda, and finding (sic.), and you have a bastardised **pseudo-science** barely worth another look. Unscrupulous people **making money** out of the latest **political bandwagon**, to justify further taxation by this dreadful government.

In the following example not only 'normal' science or ideal type of (Popperian) science is appropriated by the commentator for an anti-(normal climate) science discourse, but the

same goes for 'sceptical', that is the ideal type of Mertonian scientific scepticism is appropriated too -- by what one normally calls 'climate sceptics' to contest (climate) science.

[4] Climate research and IPCC AR4 report is a science fiction. -Fiction may have sometimes some **scientific base** to make the story more realistic. -Science is based on **verifiable facts** and **sceptical dialogue**.

In this way, two pillars of scientific thinking are seen as lacking in climate science. This is based on the belief that, as purportedly revealed by the hacked email exchanges between climate scientists at the University of East Anglia, scientists expressed concerns over frequent Freedom of Information requests, that is, had doubts about the legitimacy of these requests. This led to allegations that the scientists did not want to engage in open scientific debate, did not want to expose their data to falsification, and, in addition, wanted to suppress some data and some critical scientific voices (Pearce 2010).

The strategy of creating associations between scientists and 'corrupt' politicians is not surprising in the context of the Daily Mail and its long history of exposing political scandals, dating back to eighteenth century campaign against 'Old Corruption' (Hollis 1970). Conboy (2008), for example, notes that in UK tabloids politicians are represented almost exclusively as involved in a game of duping the public, while implying that the tabloids themselves are doing their best to rectify the situation. As our comparative analysis of keywords has shown, associations between science and fraud, and scientists and politicians, were present in earlier comments before 'climategate', and were also observed in other studies of social media (Nerlich and Koteyko 2009). However, the topics and discursive strategies employed by tabloid readers in the 2010 corpus indicate that the East Anglia controversy might have allowed the commentators to become more assured in their assertions about this link; and to speak as if they now had proof for their views that the science of climate change was indeed 'just a money-making scam'. In other words, the controversy became a rhetorical stepping stone that allowed those tabloid readers who are opposed to the idea of man-made global warming to launch more provocatively phrased comments on the issue.



Figure 4: Concordances of 'scien*' in the context of 'so-called'

The strategy based on supporting the main tenets of science appears to be novel in the context of online reader comments. It is based on the appropriation of the arguments

typically used by scientists themselves against 'climate denialism' and draws on one aspect of the scientific method often referred to as 'organised scepticism', as scientists aim to disprove ideas put forward by fellow researchers (Merton 1942). Instead of merely following the well-trodden track of climate deniers who frequently invoke uncertainty or certainty (Nerlich 2010; Reisigl and Wodak 2009) as a justification for inaction (the 'science is not settled' or the 'science is a religion/dogma/tyranny' arguments), some of the commentators in our corpus took a step further to construct two groups of scientists – the 'real' and 'proper' ones and the 'agw scientists', and differentiate between them to support their positions. This is why in our data we have a frequent collocation 'so-called scientists', whereas Reisigl and Wodak (2009, 113) commented on the use of the phrase 'so-called global warming'. Such differentiation functions as a mitigation strategy and merits a more detailed discussion.

Since 1970 studies in the fields of history and sociology of science have charted the rise of the epistemic authority of science, which despite increasing contestations, still holds a prominent hold on public debate (Porter 1995). This is the authority referred to in those comments that challenge climate sceptic arguments, made either in the news article or in previous posts, as in the following example: [5] 'Given the choice of believing a book written by a fruitcake journo like Booker or climate scientists who actually know what they're talking about, I know which choice to make. And it's not Booker.' In this context then, demonstration of scientific knowledge affords higher credibility to the speaker/writer, which is a popular legitimisation strategy in computer-mediated communication, where one's professional or any other status indicators are not visible (Galegher et al 1998; Armstrong et al in press). Conversely, a straightforward denial of scientific authority is potentially risky since the speaker/writer can be accused of irrationality and subjectivity. It is therefore not surprising that in this context of peer-to-peer interaction, users of the above mitigation strategy present themselves not as denouncing science per se, only as questioning some flawed 'subset' of it, and in doing so, they appear to subscribe to the Mertonian norms of disinterestedness and scepticism (Merton 1942), as in examples [1] and [4] above.

The impartiality and scepticism claimed in the subscription to the norms of scientific debate are only superficial however, as commentators do not engage in the scientific debate by providing proof in the form of data or reasoning; they merely evoke science. Such evocation functions as a disclaimer similar to denial statements such as 'I'm not a racist, but . . .' studied by Van Dijk (1992), which are used to express racist views covertly in order maintain a positive face because of strict social norms against discrimination and racism. Following the same line of argument we suggest that scholars of climate change communication need to be mindful of mitigation strategies like this since there is a risk that the language of climate denial can become more covert, turning into a seemingly more socially acceptable version where anti-scientific bias is carefully mitigated.

Comment ranking: adding another dimension to the popularity of arguments

Analysis of textual content can bring valuable insights into the popular topics and framing

strategies; however, using only message-level analysis does not address the dynamics of online contributions. In case of reader comments, this means that posts can gain prominence in part by attracting positive or negative ratings as they are automatically moved towards the top of the page. The platform from which our data was harvested allows rating of published comments as positive or negative (support or disagreement), a process known as 'social tagging' (Gupta et al, 2011). The comment [3] cited in the preceding section, for example, appears at the top of comments positively ranked in terms of other readers' votes (it achieved a positive rating of 1235, one of the highest ratings attracted by comments in the corpus). This tendency of favouring posts expressing negative attitudes holds across the first 50 news articles that attracted comments on climate change in 2010 and adds another dimension to the popularity of particular arguments in texts under study.

Although an in-depth study into the mechanisms behind readers' actual uses of such ranking is required, this tendency to support negative and provocatively phrased comments merits further discussion as it raises questions about the role of such rating. First, readers may reply more frequently to the first and prominent post which may affect the topics discussed. This also may affect the content of posts indirectly as online posters, found to be primarily concerned with issues of self-representation in studies of social media (Papacharissi 2009), will strive to produce increasingly more provocative or authoritative comments to gain distinction in the crowdedness of online spaces (Myers 2010). This might be the case in our sample as some of the more provocatively phrased posts and/or posts displaying knowledge of 'real' science appeared to be among the highest ranked. Second, future research needs to explore and take into account the visibility of particularly high rated comments and how it may affect the popularity of the comments among the online audience in general (it can attract additional 'passive readers' or lurkers who engage with the online content through supporting or rejecting specific posts but can also re-post it on other online platforms) and among newspaper editors and journalists (high ratings may be taken as an indicator of reader preferences).

Conclusions

In this article we set out to demonstrate how corpus-assisted approach can be employed for an inductive, bottom-up study to uncover the tabloid readers' discourses on climate change, as well as how they changed over time and potentially in response to 'climategate' -- defined by Holliman as 'a scientific and political news story that explored the validity and reliability of symbolically significant aspects of climate science' (2011, 833). In line with previous corpus-assisted studies of print media discourses (Baker et al, 2008; Partington, 2010), the benefits of using corpus-based methodology to analyse online comments lie in detecting emerging patterns via keywords, pinpointing areas that warrant further in-depth investigations (and therefore formulating strategies for downsampling) via collocational networks, and facilitating qualitative contextualised analysis. Below we discuss each of these methodological advantages from the perspective of a socio-cultural research on climate change.

Using keywords, the corpus-assisted approach allowed for the comparison of patterns in readers' comments during two different time periods, separated by the controversy that

still lingers in the media spotlight (after the release of a new batch of CRU emails in 2011, dubbed 'climategate, the sequel', the release of documents from an climate sceptic think tank, dubbed 'denialgate' in early 2012 and so on). Analysis of collocations allowed us to reveal repeated expressed meanings showing that comments from both periods contained pejorative use of the term 'scientists'. However, whereas in 2009 the issues of scientific data and the nature of scientific enquiry were marginalised, in 2010 these topics took centre stage. Although perhaps not surprising, these findings give empirical evidence to support our intuition regarding the influence of 'climategate' on discussions of science in tabloid reader comments, confirming theoretical speculation by other researchers about its rhetorical 'power' in the digitally mediated sphere (Holliman 2011). Following Partington's (2003, 12) argument: 'At the simplest level, corpus technology helps find other examples of a phenomenon one has already noted. At the other extreme, it reveals patterns of use previously unthought of. In between, it can reinforce, refute or revise a researcher's intuition and show them why and how much their suspicions were grounded'.

The data visualisation tool of concordances complemented by statistics on comment rating allowed us to investigate how online tabloid readers make their opinions believable and reasonable (Billig 1991), paying attention to the content and linguistic characteristics of lay talk as well as to the specific parameters of the web-based interactional context. Some of the delegitimisation strategies explored in this article, such as association between science and politics, or emphasis on scientific uncertainty are typical of climate sceptic logic (Nerlich 2010; Reisigl and Wodak 2009). Their reproduction in readers' comments indicates that larger political and media discourses may to a certain extent 'pre-formulate' beliefs that later gain popular currency (Van Dijk 1992, 88). If the popularity of such arguments indeed is or becomes widespread they can contribute to a process that may make people unsure about what science is, how it works and how it contributes to policies around climate change mitigation.

At the same time, other rhetorical strategies such as intensification and mitigation, but also definitional battles evident in the use of nomination strategies demonstrate that the climate sceptic positions formulated in the reader comments are not simply a reproduction of beliefs and opinions expressed in the tabloid news article. They are also a response to the arguments expressed by earlier readers, as well as an anticipation of replies that can challenge their position. Such engagement with previously posted content is not extensive, especially when compared to blogs or popular discussion fora, and is best described within the parameters of agonistic pluralism (Papacharissi 2009) rather than the Habermasian ideal of democratic deliberation. However, this aspect of online reader comments is valuable for what it can reveal about how a climate denial stance is adopted or resisted in peer-to-peer interaction couched in the 'language of common sense' (Tolson 2001, 26).

Earlier studies have documented positive and negative outcomes of 'climategate', such as throwing into sharper focus the debate about the role of science in policy making process, about openness in science, and data sharing (Holliman 2011) on the one hand, and vilification of scientists (see e.g. Mann 2012), and erosion of trust in science on the

other. A less acknowledged outcome, however, is a realisation that to a certain extent 'climategate' caught natural and social scientists unprepared -- both to the possibility that a politically conservative movement can represent a serious challenge to the dominant institutional logic supporting anthropogenic global warming (Hoffman 2011), and to the role of social media in galvanising it. Although some emerging sociological analyses have since started to examine the 'top-down' frames in the print media and in talk by prominent climate sceptics and deniers (Hoffman 2011; McCright and Dunlap 2010), very little attention is still being paid to lay discourses in general, and Internet-based discourses in particular. How are climate sceptic positions articulated in the variety of social media platforms? What are the discursive elements that sustain or challenge these positions? It is hoped that the approach presented in this paper will enable social scientists to address these questions, and analyse not only the full portfolio of cultural frames and rhetorical strategies recruited in climate sceptic discourses, but also how they are reproduced or contested in web-based platforms and what underlies their popularity in this digital landscape.

References

- Armstrong N Koteyko N Powell J** in press 'Oh dear should I really be saying that on here?' Issues of identity and authority in an online diabetes community *Health*
- Baker P Gabrielatos C Khosravinik M Krzyzanowski M McEnery T Wodak R** 2008 A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK Press' *Discourse and Society* 19 3, 273–306
- Bhatia V Flowerdew and R Jones** eds 2008 *Advances in Discourse Studies* London Routledge
- Benford R and D Snow** 2000 Framing processes and social movements, An overview and assessment *Annual Review of Sociology*, 611-639
- Billig M** 1991 *Ideology and Opinions, Studies in Rhetorical Psychology* London Sage
- Boykoff M** 2008 The cultural politics of climate change discourse in UK tabloids *Political Geography* 27, 5, 549–569
- Carvalho A** 2007 Ideological cultures and media discourses on scientific knowledge, re-reading news on climate change *Public Understanding of Science* 16 2, 223–243
- Conboy M** 2008 *Tabloid Britain, Constructing A Community Through Language*, Abingdon, Routledge
- Galegher J Sproull L and S Kiesler** 1998 Legitimacy Authority and Community in Electronic Support Groups *Written Communication* 15, 493-530
- Gupta M Li R Yin Z and J Han** 2011 Survey on Social Tagging Techniques in **C Aggarwal** ed *Social Network Data Analytics* New York, Springer
- Herring S** 2004 Computer-mediated discourse analysis, An approach to researching online behavior in **S Barab R Kling and J Gray** Eds *Designing for virtual communities in the service of learning* New York, Cambridge University Press 338-376
- Hoffman A** 2011 Talking Past Each Other? Cultural Framing of Skeptical and Convinced Logics in the Climate Change Debate *Organisation and Environment* 24 1, 3-33
- Holliman R** 2011 Advocacy in the tail: Exploring the implications of 'climategate' for science journalism and public debate in the digital age *Journalism* 127 832–846
- Hulme M** 2010 16 November The year climate science was redefined *The*

Guardian Accessed 12 February 2012

(www.guardian.co.uk/environment/2010/nov/15/year-climate-science-was-redefined)

Hutchby I 2001 Technologies Texts and Affordances *Sociology* 35 2 441-456

Jaspal R Nerlich B and Koteyko N (under review) Contesting science by appealing to its norms: Readers discuss climate science in *The Daily Mail*

Koller V and Mautner G 2004 Computer applications in critical discourse analysis in **C Coffin A Hewings and K O'Halloran** eds *Applying English Grammar: Corpus and functional approaches* pp 216-228, London Arnold

Koteyko N 2012 Managing carbon emissions: A discursive presentation of 'market-driven sustainability' in the British media *Language and Communication* 32, 24–35

Koteyko N 2010 Mining the Internet for linguistic and social data: An analysis of 'carbon compounds' in web feeds *Discourse & Society* 21 6, 655-674

Koteyko N 2007 A diachronic approach to meaning in corpus linguistics: the use of English loanwords in Russian opposition discourse *Corpora* 3, 65-95.

Killingsworth M and Palmer J 1992 *Ecospeak, Rhetoric and environmental politics* Carbondale IL, Southern Illinois University Press

Lamerichs J and H te Molder 2003 Computer Mediated Communication, From a cognitive to a discursive model *New Media & Society* 54 451-473

Laslo E Baram-Tsabari A and Lewenstein B V 2011 A growth medium for the message, Online science journalism affordances for exploring public discourse of science and ethics *Journalism* 12 7 847-870

Lederbogen U and J Trebbe 2003 Promoting science on the web *Science Communication* 24, 333-352

Leiserowitz A Maibach E Roser-Renouf C Smith N and Dawson E in press Climategate public opinion and the loss of trust *American Behavioral Scientist*

McCright A and R Dunlap 2010 "Anti-reflexivity, The American conservative movement's success in undermining climate science and policy" *Theory Culture & Society* 272-3, 100-133

Merton R K 1942 Science and technology in a democratic order *Journal of Legal and Political Sociology* 11/2 115-126

Moscovici S 1988 Notes towards a description of social representations *European Journal of Social Psychology* 18 211-250

Myers G 2010 Stance-taking and public discussion in blogs *Critical Discourse Studies* 7 4, 263-275

Nerlich B 2010 'Climategate', Paradoxical metaphors and political parties *Environmental Values* 19, 419-442

Papacharissi Z 2009 The virtual sphere 20, The internet the public sphere and beyond In, Chadwick A and Howard P eds *Handbook of Internet Politics* London, Routledge 230–245

Partington A 2003 *The Linguistics of Political Argument* London, Routledge

Partington A 2010 Modern Diachronic Corpus-Assisted Discourse Studies MD-CADS on UK newspapers, an overview of the project *Corpora* 52 83-108

Pearce F 2010 9 February How the 'climategate' scandal is bogus and based on climate sceptics' lies Retrieved on 1 March 2012 from

<http://www.guardian.co.uk/environment/2010/feb/09/climategate-bogus-sceptics-lies>

Potthast M and S Becker 2010 Opinion Summarization of Web Comments *Advances in*

Information Retrieval 668-669

Reisigl M and Wodak R 2001 *Discourse and Discrimination* London, Routledge

Reisigl M and Wodak R 2009 The Discourse-Historical Approach in **R Wodak and M Meyer** eds *Methods of critical discourse analysis* 2nd ed pp 87-121 London, SAGE

Reser J Pidgeon N Spence A Bradley G A and Glendon I 2011 Public Risk Perceptions Understandings and Responses to Climate Change in Australia and Great Britain, Interim Report

Porter T 1995 *Trust in Numbers The Pursuit of Objectivity in Science and Public Life* Princeton NJ, Princeton University Press

Richardson J and J Stanyer 2011 Reader opinion in the digital age, Tabloid and broadsheet newspaper websites and the exercise of political voice *Journalism* 128, 983–1003

Scott M 2011 WordSmith Tools version 6 Liverpool, Lexical Analysis Software

Stubbs M 2001 *Words and Phrases, Corpus Studies of Lexical Semantics* Oxford, Blackwell

Tolson A 2001 Television talk shows, discourse performance spectacle Mahwah NJ/London, Lawrence Erlbaum

Van Dijk T 1992 Discourse and the Denial of Racism *Discourse & Society* 31, 87–118

[1] Our second corpus is much smaller in size, whereas for the purposes of keyword comparison the corpora should ideally contain a roughly equivalent number of words. However, we could not supplement the data with comments from earlier years as no comments were published on the topic up until late 2008.