

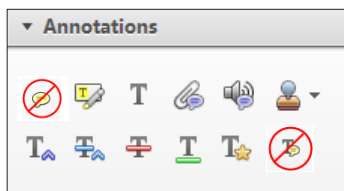
Page Proof Instructions and Queries

Journal Title: Discourse & Communication
Article Number: 568544

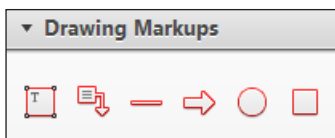
Greetings, and thank you for publishing with SAGE. We have prepared this page proof for your review. Please respond to each of the below queries by digitally marking this PDF using Adobe Reader.

Click “Comment” in the upper right corner of Adobe Reader to access the mark-up tools as follows:

For textual edits, please use the “Annotations” tools. Please refrain from using the two tools crossed out below, as data loss can occur when using these tools.



For formatting requests, questions, or other complicated changes, please insert a comment using “Drawing Markups.”



Detailed annotation guidelines can be viewed at: <http://www.sagepub.com/repository/binaries/pdfs/AnnotationGuidelines.pdf>
 Adobe Reader can be downloaded (free) at: <http://www.adobe.com/products/reader.html>.

No.	Query
	Please note that the article appears black and white in print (color in online-only). Please check Figures 1–7 that appear in color.
	Please review the entire document for typographical errors, mathematical errors and any other necessary corrections, and check headings, tables and figures.
	Please ensure that you have obtained and enclosed all necessary permissions for the reproduction of artistic works (e.g. illustrations, photographs, charts, maps, other visual material, etc.) not owned by yourself. Please refer to your publishing agreement for further information.
	Please note that this proof represents your final opportunity to review your article prior to publication, so please do send all of your changes now.
1	Please confirm that all author information, including names, affiliations, sequence and contact details, is correct.
2	Please check whether the expansion inserted for ‘CDA’ is correct.
3	Please check whether the inserted text citation for ‘LexisNexis (2014)’ in the sentence ‘Articles were collected ...’ is correct.
4	Please confirm that the funding statement is accurate.
5	Please provide the page range for ‘Charteris-Black (2014)’.



The Edward Snowden affair: A corpus study of the British press

Discourse & Communication

1–22

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1750481314568544

dcm.sagepub.com



Jens Branum and Jonathan Charteris-Black

University of the West of England, UK

[AQ: 1]

Abstract

Keyword analysis is used to compare the reporting strategies of three major UK newspapers on the topic of Edward Snowden and state surveillance. Differences are identified in the reporting strategies of *The Guardian*, *Daily Mail* and *The Sun* that provide insight into the ideology of the British press. There is significant variation in the style, content and stances of each newspaper towards state surveillance, as well as clear evidence of ideology within each paper: *The Guardian* is critical of surveillance and defensive of its decision to publish classified documents; the *Daily Mail* focuses heavily on personalisation and the personal life of Edward Snowden; and *The Sun* perpetuates an us/them ideology that is highly supportive of state surveillance. These differences are motivated by each newspaper's ideology, news values and audience considerations and demonstrate how British newspapers offer radically different perspectives on the same events.

Keywords

Critical discourse analysis, comparative keyword analysis, corpus, Edward Snowden, ideology, press, surveillance **[AQ: 2]**

Introduction

The Edward Snowden affair concerns a Central Intelligence Agency (CIA) contractor who released thousands of classified documents to selected journalists. The affair began at the start of June 2013 when *The Guardian* published articles stating that the American National Security Agency (NSA) and the British Government

Corresponding author:

Jonathan Charteris-Black, ACE (English and Linguistics), University of the West of England, S-block, Frenchay Campus, Coldharbour Lane, Bristol BS16, UK.

Email: Jonathan.Charteris-Black@uwe.ac.uk

Communications Headquarters (GCHQ) were conducting widespread secret surveillance of the public. The identity of the 'whistleblower' was revealed a few days later, and the story became worldwide news. It had a substantial political impact and opened up debates about public awareness of surveillance, exposing a conflict between the rights of citizens to privacy and the security threat posed by the release of classified documents. Fowler (1991: 1) argues that the contents of newspapers do not represent facts about the world but sets of beliefs, values and ideologies that are linguistically constructed. This study sets out to compare the linguistic choices of three British newspapers in their reporting of the Edward Snowden affair in order to identify wider ideological differences between newspapers. First, we review relevant literature and describe the methods of a corpus-based enquiry into press reporting of the Edward Snowden affair. Then, after presenting the findings of the keyword analysis, we discuss what our results reveal about the ideology of the British press.

News values, audience and ideology

Fowler (1991: 10–13) states that the news is not a plain retelling of events but a social construction that is engineered through a complex process of selection. He challenges the notion of impartiality within the press and implies that due to social, economic and political factors, news organisations always report stories from a certain angle or perspective. Events that are reported by the media have no intrinsic newsworthiness but only become news after they have been selected and included within reports. This process of selectivity is strongly influenced by the news values of a particular media organisation.

Richardson (2007: 91) describes news values as the criteria employed by journalists to evaluate potential stories and establish their suitability for coverage. Harcup and O'Neill (2001: 261–279) suggest a consistent set of news values used by journalists when they select stories, such as 'the power elite, celebrity, entertainment, surprise, good news, bad news, magnitude, relevance, follow up and newspaper agenda', and although the list provides a good overall summary of the kinds of features which journalists consider when selecting news, Richardson (2007: 92–93) emphasises the importance of audience preference. He explains that news values are based on what the intended readership are believed to find interesting.

Sumpter (2000) emphasises audience considerations in the selection of news and describes how editors create imaginary audiences to simulate reactions to different stories: 'these socially constructed readers often resembled the interests and demographics of the people in the newsroom' (p. 338). The use of an imagined audience to determine the selection of news illustrates how the practices of the press shape the final outcome of reporting. DeWerth-Pallmeyer (1997) describes the preconceived ideas journalists have towards audience expectations of news as 'audience images' (pp. 1–10). He states that these perceptions of audience 'shape' the work of journalists at all levels and suggests that an audience can be considered a form of 'social construction'. Cotter (2001) also emphasises the importance of the relationship between audience and journalist and proposes that examining the 'intentions of journalists in relation to their audience' is vital to understanding how news is selected and constructed (pp. 421–429).

The notion of ideology is also relevant to the implied audiences that newspapers consider when composing and selecting stories. Van Dijk (2001) describes ideology as 'beliefs shared by groups', including attitudes, norms and values which 'form the basis' of their belief systems (p. 12). He asserts that ideologies are not inherently negative but are dependent on the ways in which they are socially practised: positive representations can 'organise dissidence and opposition' to oppressive regimes, whereas negative ones can be used to 'establish and maintain social dominance'. Purvis and Hunt (1993: 474) relate ideology to human attempts to comprehend the world around them and demonstrate how this process forms an awareness of the social world. They highlight the importance of institutions in spreading these belief systems, which grants them significant influence over how individuals perceive the world. This links to the concept of an implied reader and suggests that newspapers contribute to the creation and maintenance of ideologies that fit their agendas through 'reporting stories in a way that is designed to evoke one particular response, thus establishing a set of shared values' (Reah, 2002: 40). This presents a view of the press as a group of organisations with the power to influence the opinions of social groups and form their views and beliefs.

Bednarek and Caple (2014) discuss the 'ideological aspects of news values' and examine how news values are 'discursively constructed', created and reinforced in journalistic texts (p. 137). They distinguish between a 'cognitive' and 'discursive' view of news values and suggest that from a discursive perspective news values are constructed through discourse itself and that certain aspects of a story are highlighted by the way events are presented. Cotter (2010) also comments on the link between ideology and news values, explaining that news values 'are one of the most important practise-based and ideological factors' in shaping and creating news (p. 67). This supports the view that news values are largely influenced and crafted by the discursive practices of journalists. As Richardson (2007) notes,

... our assessment of the 'meaning' of a text is often affected by our judgement of who produced it ... Similarly, the production of texts (and the encoding of textual meaning specifically) always has at least one eye on the imagined or target consumer and the kind of texts that they prefer to read. (p. 41)

This demonstrates the complex interactive relationship between the journalist and his readership, and the set of news values and potential ideologies involved in the construction of news. In order to examine news from a critical linguistic perspective, it is important to consider these relationships in the context of specific linguistic choices made by individual journalists to discover how language can 'work subliminally in the newspapers' ideological practise of representation' (Fowler, 1991: 5).

With the above considerations of news values, audience, ideology and language choice in mind, this study sets out to address the following research questions:

1. What does corpus analysis reveal about the differences between British newspapers in their reporting strategies of the Edward Snowden affair?
2. What does corpus analysis reveal about the ideology of the British press?

Method

Baker (2006) discusses the advantages of corpora in discourse analysis, explaining how the ‘cognitive biases’ of the researcher can be reduced through the use of corpora. This is achieved by using computer software to examine a large selection of texts to identify ‘trends and patterns’ instead of selectively choosing particular texts that suit a particular authorial narrative. He points out several other advantages of corpus methods, such as how corpora allow ‘numerous supporting examples’ of language use to be identified, which strengthens claims made by the researcher about how language is employed. Our methodology follows a now established framework for using corpus linguistic methods to undertake critical studies of discourse (Bednarek and Caple, 2014; Hunter and MacDonald, 2013; Kim, 2014).

‘Comparative keyword analysis’ offers a further refinement of traditional corpus methodologies (Seale and Charteris-Black, 2010); traditional methods employ a large ‘reference’ corpus that is intended to be representative of language use in general; however, comparative keyword analysis replaces the reference corpus with a ‘discourse reference corpus’ by selecting *only* language that shares a particular purpose – thereby controlling for the variable of discourse. Charteris-Black (2012, 2014) demonstrates how keywords – those that occur statistically more frequently in a research corpus – can be identified by comparing the speeches of particular politicians (Tony Blair and David Cameron) with a discourse reference corpus of political speeches. This study adapts this approach to the discourse of press reporting. Three distinct research corpora, each comprising press reports from *The Guardian*, *The Sun* and the *Daily Mail*, respectively, are compared with a discourse reference corpus comprising three tabloid newspapers (the *Daily Star*, the *Daily Mirror*, *The Sun*), three middle-market tabloid newspapers (*Daily Mail*, *Daily Express*, *London Evening Standard*) and two broadsheet newspapers (*The Guardian*, *The Times*). Articles were collected from the websites of each newspaper, with the exception of *The Sun* and *The Times*, which were gathered through the news database (LexisNexis, 2014). **[AQ: 3]**

A search term and a time frame were applied across the full sample. For each of the eight newspapers, ‘Snowden’ was used as a search term for the period 1 June 2013 until 17 November 2013; the articles that were relevant to the Edward Snowden affair were then included. *The Guardian* did not have a search tool which allowed searches to be refined in this way, but had an entire section of the website, ‘The NSA files’ which contained a comprehensive list of articles related to the story in a timeline format. Table 1 summarises the contents of the research corpora (asterisked) and the discourse reference corpus:

While all the newspapers covered the story in some depth (apart from the *Daily Star*), the *Daily Mail* and *The Guardian* newspapers ran considerably more stories on Edward Snowden (accounting for 37.8% and 36.7% of the whole corpus, respectively).

The corpus analysis software AntConc (Anthony, 2011) was used to undertake a comparative keyword analysis (using chi-square tests) of the asterisked sections of the corpus (the research corpus) with the discourse reference corpus.¹

The software generated a list of the most frequently used words in each research corpus and compared these against a list of the most frequent words in the discourse

Table 1. Overview of discourse reference corpus.

Newspaper corpus	Number of words	Number of articles
<i>Daily Mail</i> *	303,552	298
<i>The Guardian</i> *	294,891	341
<i>Evening Standard</i>	66,939	131
<i>The Times</i>	50,947	101
<i>Daily Express</i>	34,405	82
<i>Daily Mirror</i>	30,348	72
<i>The Sun</i> *	20,693	111
<i>Daily Star</i>	1349	9
Total	803,124	1145

reference corpus. This creates a keyword list: a list of the statistically more frequent words which can subsequently be used to ‘form the basis of a qualitative investigation’ (Charteris-Black, 2014: 541). The keywords that formed the basis of our study were selected and evaluated qualitatively based on their potential to answer the research questions. Charteris-Black (2014: 544) notes that statistical significance is not always the best way of determining the value of a keyword since content keywords may be less insightful than other lower frequency keywords. Higher frequency keywords may only relate to the basic content of the story without revealing much about reporting strategy or authorial ideology. In this study, such discursively uninteresting content words include *officials*, *program*, *government* and proper names, such as *Russia*, *Edward*, *United States*.² However, lower frequency keywords may be more insightful; for example, *Novotel* is a lower frequency keyword for the *Daily Mail* (keyness=7.56); however, when combined with another *Daily Mail* keyword *hotel* (keyness 32.72), it shows how the reporting strategy of the *Daily Mail* emphasises the luxury of Snowden’s living conditions while in Hong Kong. Therefore, patterns of keywords reveal underlying authorial perspectives when they are semantically grouped, irrespective of whether they are lower or higher frequency.

The keyword lists are ordered by the software in terms of their keyness: a number calculated using a chi-squared test which represents the extent to which a word has a statistically significant frequency difference in the target corpus when compared to the discourse corpus. The top 150 words were chosen and examined individually with the assistance of the tools built into the software, which provide many ways of evaluating keyword significance. The collocation tool lists words which commonly occur with the keyword within certain parameters, for example, four words to the left or right of the word, and can assist the researcher in detecting patterns in the use of that keyword. The keyword in context (KWIC) display provides a list of all the uses of a keyword within its original context. This allows the researcher to examine the surrounding text for evidence of the underlying reporting strategy. For example, within *The Guardian* corpus, *reporting* appears 77 times, with a keyness of 10.32, showing that it is used statistically more often than the other newspapers. The cluster tool identifies groups of words that occur most frequently with the keyword; in the case of the keyword *reporting*, the software

Hit	KWIC
52	at looking at the issues raised by the Guardian's reporting, namely: "Is government accessing the data for t
53	n-chief said yesterday. Defending the Guardian's reporting of the Snowden files in the wake of criticism mad
54	quarters in Fort Meade, Maryland. The Guardian's reporting on the National Security Agency's surveillance ac
55	ropagated a series of myths" about the Guardian's reporting of the Snowden documents. "When responsible journ
56	of the public, have all said that the Guardian's reporting on this story is important for democracy." She a
57	Andrew Parker, for suggesting that the Guardian's reporting of the NSA files was a "gift" to terrorists. Dom
58	Office lawyer for suggesting that the Guardian's reporting of the National Security Agency files has provide
59	n-chief of the Guardian, defended the newspaper's reporting of files leaked by Snowden that highlighted formi
60	itimate inquiry on grounds of national security reporting, not document dumps proved spurious or self-ser
61	cism Former Foreign Office lawyer says claim that reporting on reach of GCHQ and NSA breaches national securi
62	ardian - for reporting on the Snowden files. That reporting has so far led to a Presidential review and three

Figure 1. 'Reporting' concordance sample from *The Guardian*, demonstrating the cluster.

identified *The Guardian's reporting* as a cluster occurring seven times. A dispersion plot can be used to show the spread of this cluster throughout the corpus to determine whether it occurs in one particular article or is spread evenly. In this case, the cluster occurred in five different articles. These groups of words are often a good place to start analysing because they show repeated patterns of discourse; examining these concordances in context can then lead to insights into how the keyword is used. The keyword *reporting*, without further examination in this way, may have been categorised as a neutral word related to the process of news making and therefore not indicative of any further purpose (Figure 1).

When examined in context, 55 of the 77 uses of *reporting* relate to the importance of the story and its need to be shared with the public, through defending *The Guardian's* decision to publish the story, responding to criticism or implying other media outlets are not covering the story in as much detail. This suggests a reporting strategy in which *The Guardian* seeks to justify its coverage and implies the prevalence of the word *reporting* may be a result of this strategy. For example, the wider context of line 57 uses the voice of a politician to defend *The Guardian* against criticisms it has received:

A conservative MP and former Foreign Office lawyer has criticised the MI5 director general, Andrew Parker, for suggesting that the Guardian's reporting of the NSA files was a 'gift' to terrorists. ('Tory MP Dominic Raab Defends *Guardian* Against MI5 Criticism', *The Guardian*, 31 October 2013)

Here, we see how the very fact of reporting a story was in itself controversial, with individuals positioning themselves in relation to major ideologies relating to tensions between concerns for freedom of the press and for national security.

This demonstrates how a reporting strategy was identified by the use of the tools in the software in conjunction with the qualitative judgement of the researcher: the cluster tool identified *The Guardian's reporting*, those concordances were then expanded and examined, and the potential theme of defence against criticism was considered in relation to each of the other 76 concordances of the keyword. We found that, as with Baker (2014: 198), it was only necessary to search an expanded context when the concordance line alone was insufficient to identify a particular reporting strategy. When this was the case, it was necessary at times to read a full paragraph of text.

Table 2. *The Guardian* keywords.

Reporting strategy/stance	Semantic/conceptual grouping	Keywords
News values		<i>data, digital, internet, metadata, privacy, protection, surveillance, technology</i>
Criticism of surveillance	Insidious surveillance	<i>data-mining, personal, power, target(s), Tor, you</i>
	Magnitude of surveillance	<i>all, bn, bulk, collection, many, million, much, significant, scale, too</i>
Justification of reporting	Legality	<i>Act, amendment(s), bill, constitutional, FISA, law, Patriot, section</i>
	Morality	<i>approval, blind, citizens, challenge, debate, democracy, freedom, fundamental, issues, legitimate, liberty, openness, oversight, public, reform, trust</i>
	Public support	<i>editorial, editors, journalism, our, press, reporting, we</i>

In this way, we detected patterns of semantically related keywords, which allowed us to code such groups of keywords according to how they provide insight into the style and reporting strategy of each paper. Although this strategy enabled us to classify groups of keywords together, it is important to note that it was not always a simple process since not every concordance of every keyword could be examined in detail or contained ambiguous examples which were difficult to categorise concretely. The following section shows the reporting strategies for each of the newspapers, based on this classification.

Findings

The Guardian

The Guardian made the original decision to publish classified information, and the keyword analysis reflects its criticism of state surveillance and defence of its choice to report the story. Table 2 shows the 57 keywords and phrases included in the analysis and their classification by semantic/conceptual relatedness and by reporting strategy/stance.

News values: Surveillance and the right to report

Many of the keywords are content words that provide insight into *The Guardian*’s news values as well as the topics of the stories – such as when words relate to either the surveillance itself: *data, protection, privacy* and *surveillance* or the technical aspects of surveillance: *technology, internet, digital* and *metadata*. A recurring theme was a portrayal of government surveillance as immoral, unjustified, and widespread. This is achieved through an emphasis on both the insidious nature of surveillance and the extent to which it took place. The keyword *Tor* refers to attempts by the NSA to compromise the browsing tool *Tor*, which ensures that Internet users remain anonymous. Within the

Hit	KWIC
3	of the NSA repeatedly developing attacks against Tor. That appears to have been a frustrating task. "
4	or users they spy on." The NSA's attacks against Tor included targeting security holes in the Firefox
5	task. "We will never be able to de-anonymise all Tor users all the time," said "Tor Stinks", an NSA pr
6	powerful anonymity service, and the fact that all Tor users look alike on the internet, makes it easy t
7	chneider recently made anonymisation tools such as Tor the first step in his advice on "how to remain se
8	e National Security Agency. The work of attacking Tor is done by the NSA's application vulnerabilities
9	to force change and improve security. Attacking Tor: how the NSA targets users' online anonymity Secr
10	cement co-operation In the past, the team behind Tor has responded to exactly this question, denying t
11	Who created Tor? The original technology behind Tor was developed by the US navy and has received abo
12	getting security holes in the Firefox web browser. Tor encourages users of its Tor Browser Bundle to upg
13	k to the NSA. Exploiting the Tor browser bundle Tor is a well-designed and robust anonymity tool, and

Figure 2. A sorted sample of the 'Tor' concordance.

corpus, 40 out of 89 uses refer to the NSA attacks on *Tor*, as can be seen in the sample of concordance lines (Figure 2).

Line 9 includes another related keyword *targets* and, as it occurs in a headline, shows how an entire article is dedicated to this subject. The use of *targets* suggests a deliberate action with military connotations and creates a representation of the NSA as sinister and powerful.

Another negative depiction of state surveillance focuses on power abuse, a theme which is reinforced by the keyword *power*. It is frequently used to refer to the uncontrolled power wielded by the state (Figure 3).

Line 7 contains a direct assertion that security agencies are using surveillance as a means of sustaining their own power, a sentiment that correlates with the other examples in the concordance. 'Abuses of *power*' are mentioned in three of the concordance lines; for example, line 13 reads,

It's the stuff of conspiracy theorist fantasies. But these abuses of power are real and are playing out on the front pages of America's papers every day. ('I Have Watched Barack Obama Transform into the Security President', *The Guardian*, 15 June 2013)

This serves to highlight the power abuse itself and to distance *The Guardian* from discredited and negatively perceived 'conspiracy theorists' by suggesting the evidence of government corruption can be clearly seen in the publication of NSA documents and their discussion in the media. These strategies portray surveillance as an act designed to protect the powerful and control citizens in an unethical and abusive way.

A set of keywords related to the magnitude of surveillance emphasises the scale on which surveillance has taken place. This is done through highlighting the number of people and countries affected, the money involved and the volume of data collected; for example, *bulk collection* implies that collection is not carried out selectively but in large, indiscriminate quantities. *Scale* (91 out of 101 uses) is also frequently used in these contexts, but has the additional property of amplifying the impact of the story itself, as shown by Line 1 in Figure 4 which describes the story itself as a 'huge scandal', suggesting it is *significant* and newsworthy.

Collocates also provide evidence for the focus on the level of surveillance: The two most frequent collocates of the keyword *too* were the keyword *much* (21 out of 186) and

Hit	KWIC
7	our security The NSA-GCHQ machine is about global power, not protecting its citizens. US and British inte
8	enigrated quality of privacy, but about political power. What the Guardian-New York Times stories of last
9	really need is a larger conversation about state power and the actual risks we face, with context and cl
10	y. There is a totally legitimate debate about the power of these technologies, about how you get the bala
11	lations about surveillance did not prove abuse of power, he said. "I don't think they crossed the line. I
12	are more likely to turn a blind eye" to abuses of power. He warned against the "acid turned up by the pot
13	conspiracy theorist fantasies. But these abuses of power are real and are playing out on the front pages o
14	s to give aid to murderous enemies, but abuses of power must be revealed NSA emblem One point Edward Sn
15	g that schedule 7 to the Terrorism Act 2000 gives power to the police and others to question those who ar
16	lusioned. "Unfortunately, shortly after assuming power, he closed the door on investigating systemic vio
17	ticians familiar with the complex algorithms that power the internet. It has room for a sprinkling of acc
18	rising data and use, because it allows computing power to scale up faster than Moore's Law, so long as t
19	reat. Yet because paranoia is the classic ally of power, counter-terrorism has inflated its enemies and t
20	tuation, combined with its economic and political power, comes as an opportunity and as an obligation. We
21	e whose value I have learned from the net and the power it gives any citizen to speak publicly; to find,
22	nds seurocrats unlimited money and unconstrained power. The US and Britain today are as invulnerable to

Figure 3. A sorted sample of the ‘power’ concordance.

the word far (16 out of 186), ‘too far’ in this sense is used to imply surveillance has taken place on a level beyond what is necessary and has progressed past an acceptable point.

A stance taken by *The Guardian* to justify its reporting of the story is by arguing that publishing the story was an informed ethical choice that contributes to legality, morality and has public support. A set of keywords relates to the legal aspects of surveillance. *Law, Patriot, Act, section, FISA* and *bill* all have a similar function of questioning the legality of mass surveillance or referring to specific acts and bills relevant to the story as a preliminary to challenging their validity or suggesting modifications. *Constitutional* and *amendment(s)* refer to the US Constitution, the supreme law in the country that describes the structure of government and outlines a set of amendments which should always be adhered to. The first 10 amendments are known as the ‘Bill of Rights’, and *The Guardian* frequently refers to the First Amendment, relating to freedom of speech, and the Fourth Amendment, concerning the right to be secure against searches without a warrant. Concordances of these keywords show how *The Guardian* believes that Edward Snowden, and by extension those reporting on the NSA, deserves to be protected by these fundamental rights. For example, the headline and subheading of an article reads,

America’s whistleblower persecution sullies the US constitution.

By what right is the United States exempt from the first amendment principles it demands be applied elsewhere? (*The Guardian*, 3 July 2013)

The use of these terms by *The Guardian* can be seen as a justification for publishing the documents on the grounds of freedom of speech and freedom of the press.

A further group of words are categorised as *morality*, due to their appeal to ethics, legitimacy and the perceived positive values of a progressive society. These values imply that reporting is morally justified and is done for civic benefit. *Fundamental* is used to refer to fundamental rights regarding privacy or asylum in 26 out of 36 uses; *liberty, freedom, democracy, openness* and *trust* are all examples of civic values, and *approval,*

Hit	KWIC
1	arket surveillance of the British public on such a scale amount to a huge scandal even by the standards of
2	the reaction to the Guardian's disclosures of its scale and reach certainly do. We need not be too smug h
3	ce of just one month. This is spying on a massive scale and clearly not exclusively directed at combating
4	ments have made careful decisions about both the scale and powers of organisations like MIS, proportiona
5	the elusive whistleblower Snowden. But it is the scale and reach of the NSA-GCHQ operation and the eff
6	r Edward Snowden, has reported extensively on the scale and scope of the NSA's activities, leading to leg
7	a mundane reality. Back then, we assumed that the scale and pace of growth of the internet would put it b
8	that US intercepts French phone calls on 'massive scale' Angelique Chrisafis in Paris and Sam Jones : N
9	texts to do not involve collection of data on the scale at issue in the telephony metadata collection pro
10	e way we are governed, but government made a full-scale attempt to censor their publication. Same yet aga
11	civil libertarians. He led a futile 2011 fight to scale back the Patriot Act and famously spoke for 13 st

Figure 4. A sorted sample of the 'scale' concordance.

reform, debate, oversight, public and *challenge* all dispute the fairness of surveillance. For example, a concordance of *oversight* reads,

Pressure is growing on the White House to explain whether there was effective congressional oversight of the programs revealed by Snowden. ('US Lawmakers Call for Review of Patriot Act after NSA Surveillance Revelations', *The Guardian*, 10 June 2013)

This challenges the activities of the NSA on the grounds that it is not monitored independently or approved by the public and has not been in a position of scrutiny and evaluation until the publication of the story.

This call for public debate on surveillance can be explored further through consideration of the collocates of *public* which occur one word to the right, typically in the pattern '*public interest*' (62 out of 554) and '*public debate*' (22 out of 554). These words are used by *The Guardian* to suggest its reporting is morally correct, and in the following article, parallels are drawn to other situations in which whistleblowers have spoken out against bad practice:

As nurse in the A&E department, Donnelly spent years insisting that there were serious problems at Stafford Hospital before they eventually came to light. I don't know the Snowden case, but I think that if anybody has genuine concerns that are in the public interest, then they are duty-bound to raise them. If you're a nurse, especially, it's part of your professional code of conduct to speak up if you see things that are not right. ('Edward Snowden and Whistleblowers: "The Truth Sets You Free"', *The Guardian*, 11 June 2013)

Here, the phrase '*public interest*' is used to indicate that reporting of surveillance is for the greater good in a liberal democracy that values freedom of speech. An interesting phrase which adds conviction to this stance can be found in the collocate of the keyword *blind*. Eye is the most frequent collocate (5 out of 17) and is used in the phrase '*blind eye*' to describe the attitudes of politicians, newspapers and government employees in reaction to the story. For example, the concordance 'Most papers, however, turned a blind eye' suggests that since those with the obligation to discuss the importance of the surveillance are failing to do so, it is the responsibility of *The Guardian* to ensure the public is fully informed so that a fair evaluation can be made of the need for state surveillance.

The final group of *The Guardian*'s keywords relate to the concept of 'public support'. These keywords are used in a similar way to the *reporting* example (see Figure 1) and reveal a further way that *The Guardian* defends itself against criticism by showing the support it has from members of the public and their media voices. The use of plural pronouns, such as *we* and *our*, is used in a way which subtly reinforces *The Guardian*'s reporting as a process carried out in the interests of a public group with a shared set of views, for example:

This is the world we accept if we continue to avert our eyes. And it promises to get much worse. ('The Cyber-Intelligence Complex and its Useful Idiots', *The Guardian*, 1 July 2013)

This conveys a sense that everybody is affected by issues of surveillance and that a unified effort is required to prevent the compromise of democratic values. Defence against criticism is also achieved through reference to other *editors* and *journalists* who support *The Guardian*, and by citations of *editorials* which are supportive, and responses to *editorials* that are critical, including a particularly strong criticism from the *Daily Mail*. The keyword *freedom* is the second most frequent collocate of the keyword *press* (32 out of 220 uses), and the most frequent collocates of the keyword *journalism* are 'investigative' (12 out of 83), 'good' (6 out of 83) and 'responsible' (3 out of 83). These words are used to describe the nature of *The Guardian*'s work in a positive way. Together they tell a story in which *The Guardian* justifies its reporting by appeals to morality and legality, by portraying their motives as ethical and driven by concern for the public and by emphasising public support for their reporting practices.

The Daily Mail

The keywords in the *Daily Mail* corpus contained many of the neutral words mentioned in the 'Method' section which were not considered for further analysis, leaving a final group of 38 keywords which provide insight into the news values and reporting style of the paper. These keywords can be seen in Table 3:

News values: Lifestyle and personalisation

A recurring theme reflecting the news values of the *Daily Mail* was the exotic lifestyle of Edward Snowden since he had made his revelations. Keywords such as *Novotel*, *room(s)* and *hotel* imply that Snowden was enjoying an apparently lavish holiday lifestyle based perhaps on the audience's media frame of a mystery holiday to an unknown but exotic destination. This is evident in Figure 5:

The collocations of adjectives such as *luxury* and *massive* with *hotel* attribute the qualities of wealth and prestige to Snowden's living conditions. For example, Line 68 is the headline to an entire article, which reads,

It's a hard life being on the run: Inside Edward Snowden's luxury hotel which has its own swimming pool. Turkish baths, gym and library. (*Daily Mail*, 4 July 2013)

Table 3. *Daily Mail* keywords.

Reporting strategy/stance	Semantic/conceptual grouping		Keywords
News values	Focus on Snowden’s exotic lifestyle		airport, asylum, extradite, flight, hotel, Novotel, plane, room(s), temporary, terminal, transit, unknown, where
	Focus on Snowden’s personality and personal life		anime, daughter, father, girlish, girls, he, her, him, his, Lindsay, Mills, school, son, women
Reporting style	Reporting words	Neutral	said, says, spoke, told
	Naming strategies	Propositional	according, believed, claimed fugitive, leaker, miss

Hit	KWIC
61	un, location unknown, after he fled the Hong Kong hotel where he is holed up A former whistleblower give
62	two days after Snowden checked out of a Hong Kong hotel where he was interviewed by the Guardian newspaper
63	than 70 rooms have been rented in a nearby luxury hotel for the president's security, staff and entourage
64	than 70 rooms have been rented in a nearby luxury hotel for the president's security, staff and entourage
65	special place' for them. Snowden fled to a luxury hotel Hong Kong weeks ago after he decided to release t
66	owden, last known to have checked out of a luxury hotel in the city's Kowloon district on Monday. The Sec
67	rently unknown after he checked out of his luxury hotel on Monday - though he is believed to still be in
68	being on the run: Inside Edward Snowden's luxury hotel which has its own swimming pool, Turkish baths, g
69	orhood just across the harbor from the mainland. Hotel hide-out: The Mira Hotel, pictured, in the Tsim S
70	tion about the United States Atrium: The massive hotel is thought to be home to Edward Snowden as he see
71	tion about the United States Atrium: The massive hotel is thought to be home to Edward Snowden as he see
72	use the owner wanted to sell the house. The Mira Hotel in Hong Kong, where 29-year-old former CIA employ
73	e city. But today an employee at the 5 star Mira Hotel in the Tsim Sha Tsui neighborhood said a guest by
74	e. Checked out: Today an employee at the 5* Mira Hotel in the neighborhood of Tsim Sha Tsui, said a gues

Figure 5. A sorted sample of the ‘hotel’ concordance.

Irony is used in the headline by contrasting ‘hard life’ with all the desirable features of the airport hotel Snowden stayed in, which suggests ‘being on the run’ is also pleasurable. This can also be seen as a distancing strategy, in which it is implied that Snowden is the member of an elite who can afford to stay at expensive hotels, unlike the imagined audience who are unlikely to stay in such places. This focus on his luxury lifestyle also contains a suggestion of hypocrisy since it portrays him as more concerned about his own personal comfort than the interests of the public. The article includes large pictures of the hotel that contain a short caption and passages of text that focus on Snowden and the likelihood he will gain asylum. Caple (2013: 36–38) explains how images have an important function in establishing news values within a text itself, and this is evident within the *Daily Mail* article, in which the personalisation of Snowden is emphasised by the content and language of the headline and text and foregrounded further through placement of multiple pictures of an expensive hotel. Analysis of discourse at a higher level, in this case an entire article, is an important consideration because it shows how patterns of discourse are structured beyond individual words and sentences to form an overall perspective.

Hit	KWIC
37	mainland China. mills Stunning ballerina Lindsay Mills, 28, took to her blog on Monday to detail her hea
38	to be seen. Lindsay Mills Left behind: Lindsay Mills, Snowden's girlfriend, is said to be 'barely hold
39	usic from Quentin Tarantino's 'Kill Bill' Lindsay Mills, 'My world has opened and closed all at once. Lea
40	er. Rika Rossing videoed a performance by Lindsay Mills called, somewhat aptly, 'Fish out of Water' last
41	CLOSED ALL AT ONCE': Complete posting by Lindsay Mills For those of you that know me without my super he
42	eaving me lost at sea without a compass,' Lindsay Mills writes about being suddenly left by NSA whistle b
43	eaving me lost at sea without a compass,' Lindsay Mills writes about being suddenly left by NSA whistle b
44	en was living with his girlfriend, dancer Lindsay Mills (pictured), in Hawaii until he fled He wrote that
45	9, wanted to be with stunning pole dancer Lindsay Mills for the rest of his life but took the decision to
46	Edward Snowden? Ballerina and pole dancer Lindsay Mills, 28, lived with Snowden in Hawaii She added cryp
47	r exiled lover. Ballerina and pole dancer Lindsay Mills, 28, who lived with Snowden in Hawaii before he e
48	ncluding one June 10 upload Tiny dancer: Lindsay Mills has posted numerous photos and videos on her pers
49	ed to still be in Hong Kong. Devastated: Lindsay Mills, the 28-year-old girlfriend of Edward Snowden is
50	zon, sex goddess - but a petite dreamer.' Lindsay Mills Better times: Lindsay Mills seen here with Ed Sno
51	oked not unlike Snowden's heartbroken ex Lindsay Mills on a Moscow red carpet recently. Now she may be a
52	bears a slight resemblance to Snowden ex Lindsay Mills Former Russian spy Anna Chapman has proposed mar
53	nstead fallen on his American girlfriend, Lindsay Mills, a dancer who posted partially nude photographs o
54	join him in exile. Snowden and girlfriend Lindsay Mills, were living together in Hawaii when he went publ
55	nal secrets have also left his girlfriend Lindsay Mills, 28, completely devastated. Lindsay had previous
56	to be the mother of Snowden's girlfriend, Lindsay Mills, is pictured outside her home in Annapolis, Maryl
57	m herself Edward Snowden's 'girlfriend' Lindsay Mills' sexy poledance 'I like my girlish figure that a

Figure 6. A sorted sample of the ‘Lindsay Mills’ concordance.

Another group of related keywords reveals personalisation as a major news value of the *Daily Mail*. More specifically, these words relate to his personal life, rather than his physical journey and situation. The grammatical pronouns *her*, *him*, *he* and *his* were included in this category: These are expected to be common words, but their extreme prevalence within *Daily Mail* articles compared to newspapers in the discourse reference corpus show their use is noteworthy. Closer analysis shows the majority of these pronouns relate to Snowden himself (*his*, *him*, *he*) and most of the female pronouns (*her*) to his girlfriend or other women who have a connection with him. This emphasis on Snowden himself and those related to him can be seen in family references in the keywords *father* and *daughter*, which refer to Snowden, his girlfriend Lindsay Mills and both of their fathers. Coverage of Edward Snowden’s girlfriend is frequent in the *Daily Mail* corpus, as shown in Figure 6 by the concordance of her name:

The decision to include *Lindsay Mills* is ideologically significant: Since the initial story published by *The Guardian* was about state surveillance and did not name Snowden as the source of information, stories relating to the girlfriend of the source are only distantly related to the original theme. The concordances which contain the terms ‘stunning ballerina’ (Line 37) and ‘pole dancer’ (Lines 45–47) show how sexuality, as well as personalisation, is a news value which is foregrounded. Conboy (2006) writes that tabloid content is made up of both ‘naughty innuendo’ and ‘explicit sexual material’ (pp. 29–30). While the *Daily Mail* usually appeals to the mid-market, here we find its reporting on *Lindsay Mills* follows a strategy similar to a tabloid with constant reference to her physical appearance, a trend which is mirrored by the images of her in revealing outfits or sexual poses.

The keywords *girlish*, *women* and *girls* also focus on Snowden as an individual, report on aspects of his sexuality and in some respects attempt to discredit him. For example,

women is used to refer to Snowden either ‘attracting young women’ or to receiving ‘attention from Russian women’, which follows the trend of sexualisation and evokes the schema of scandal and spy stories characteristic of the Cold War period – in actual events such as the Profumo Affair and in fictional narratives by authors such as John le Carré. *Girlish* appeared as a keyword in a concentrated section of the corpus particularly in the following *Daily Mail* article. The headline reads,

‘I like my girlish figure:’ How a geeky 18-year-old Edward Snowden bragged about video games and girls before he became the NSA leaker. (*Daily Mail*, 12 June 2013)

Richardson (2007) notes the importance of lexical choice in the construction of texts, stating that words can contain ‘value judgements’ as well as ‘connoted’ and ‘denoted meanings’ (p. 47). This is evident in the above headline, with Snowden described as ‘geeky’ and reported with the verb ‘bragged’ to suggest he is proud and vocal about sexual achievements. Richardson (2007) refers to this as ‘transformed indirect quotation’ and explains how a reporter can remove a reporting clause to replace it with a transitive action, in this case ‘bragged’, in order to alter the representation substantially (p. 104). The preposition ‘before’ is also deliberately vague since it does not state a specific time; the quote comes from an Internet blog written by Edward Snowden 11 years ago; readers may infer that Edward Snowden has written the post recently and is therefore an immature and perhaps unreliable source.

Naming strategies are the referential strategies used to describe individuals and groups with the result of conveying connoted meanings or ascribing attributes. This can be seen with the keywords *fugitive* and *leaker*: These are negative terms used to refer to Edward Snowden. This strategy is used far more frequently in *The Sun* corpus, which we will now discuss.

The Sun

The Sun corpus was by far the smallest of the three, and as a result, some of the keywords were statistically highly significant although they had a very low frequency count. Baker (2006) suggests grouping together keywords which have a low frequency count but share a similar ‘meaning or sense’ (pp. 143–144). We employed this approach to group together similar keywords that individually have a low frequency (although are still keywords) but appear more salient when considered as a ‘cumulative whole’. Therefore, each of the groupings in the following table represents a group of keywords which share a sense and demonstrate a pattern of discourse, despite their individually low frequency counts. The resulting list of 50 keywords, shown in Table 4, provides insight into the stance and reporting style of *The Sun*.

News values: Defence of surveillance rights and nationalism

The news values of *The Sun* show an informal approach to discussion of surveillance: *spooks*, *snooping*, as well as reference to the documents as *secrets* and *leaks*. This implies a negative evaluation of the publication of the documents, a pattern which can also be

Table 4. *The Sun* corpus keywords.

Reporting strategy/stance	Semantic/conceptual grouping	Keywords
News values		<i>fugitive, leak, secrets, snooping, spies, spooks</i>
Defence of surveillance	Damage caused	<i>agents, brits, distribute, endanger, lives, names, put, risk</i>
	Positive aspects of surveillance	<i>brave, listeners, safety, smears</i>
	Minimisation of significance	<i>fuss, dreary</i>
Reporting style	Naming Edward Snowden	<i>dropout, ed, ex-CIA, geek, geekish, lad, leaker, leaks, rebel, spook, traitor</i>
	<i>The Guardian</i>	<i>broadsheet, leftie, probe, wing</i>
	David Miranda	<i>mule, quiz</i>
	Government/state	<i>boss, cops, chiefs, spymaster</i>
	Lindsay Mills	<i>acrobat, beauty, lover, seminaked, sexy</i>
	Enemies	<i>fanatics, jihadis, maniacs, plotters</i>

seen in the next set of keywords which show *The Sun* is largely defensive of surveillance. These keywords have been further refined to demonstrate how *The Sun* downplays the significance of surveillance by highlighting its positive aspects and emphasising the threat to security caused by the publication of classified documents.

Conboy (2006) outlines how the ideology of nationalism is spread by tabloid newspapers and states the consistent use of terms such as ‘brits’ can be used to ‘reinforce the national community in populist form’ (pp. 46–68). Drawing on Anderson’s notion, he writes that this appeal to national community is directed towards the ‘imagined community of its readers’ and that nationalism is sustained partly through narratives that are projected through the ‘institutional support’ of national newspapers. *The Sun*’s staunch defence of state surveillance can be seen as a reflection of this pattern, and the reference to *brits* and *agents* is used patriotically to suggest the publication of NSA documents is directly threatening to their *lives*, which have been *put at risk*. For example, Line 8 of the concordance (Figure 7) reads,

Those who back the rights of hackers to leak everything about our security services – and put our brave spies’ lives at risk – have had a bad week. (‘Bad Week for Leaks’, *The Sun*, 3 November 2013)

This extract contains a presupposition that those who support the ‘hacker’ Edward Snowden are comfortable with putting the lives of British agents at risk; the statement suggests it is not possible to be concerned about both his rights and the safety of the nation at the same time. The claim that such individuals have had a ‘bad week’ serves to portray them negatively and their viewpoint as unfavourable. This and similar examples are shown in Figure 7 by the concordance of the keyword *risk*.

The other keywords in this grouping serve a similar function, by focusing on the implied threat both to the British nation as a whole and to the spies who are depicted as protectors of national security and guardians of the public.

Hit	KWIC
1	k of security material can put lives of agents at risk and give valuable information to terrorists." Mea
2	disclosure of this information puts Americans at risk." Russia yesterday suggested it could grant him p
3	disclosure of this information puts Americans at risk." Russia yesterday suggested it could grant him p
4	ntelligence is "entitled to know" who could be at risk. They are backed by a handful of Labour and Lib D
5	"put our dedicated agents, and their families at risk". PM David Cameron also backed a Commons probe, s
6	ld endanger UK national security and put lives at risk", a Government lawyer said. Mr Miranda - whose pa
7	ld endanger UK national security and put lives at risk", a Government lawyer said. Mr Miranda - whose pa
8	ty services - and put our brave spies' lives - at risk - have had a bad week. First the Brussels bureauc
9	been very damaging. They've put our operations at risk." At the same committee meeting, the head of MI5
10	about how many agents have been named and put at risk. Conservative Julian Smith insisted that in his r
11	ose who reside in allied nations have been put at risk." POOK BLAST MINISTERS have let down spies b
12	MI6, Sir Nigel Inkster said lives had been put at risk. And former GCHQ boss Sir David Omand said he fea
13	result. He added spies' lives had also been put at risk - the most serious criticism of the newspaper by
14	how many agents' lives its spy leaks have put at risk, MPs demand. A group of 28 Tories have written to
15	GIFT TO FANATICS ; SPOOKS HELPLESS ; SECURITY AT RISK AL-Qaeda have been "rubbing their hands with g
16	s life. The US claims he put national security at risk but Snowden's real crime is that he has exposed u
17	en it is right that the police act." Spy leaks 'risk lives' TERROR cops are probing after "sensitive

Figure 7. A sorted sample of the 'risk' concordance.

Another group of keywords are semantically related to a positive evaluation of surveillance. The security services are represented as stealthy but heroic agents who are *brave* and good *listeners* (rather than peeping toms) and are concerned with the *safety* of the nation. The release of documents is framed by the keyword *smears* as a spiteful attack. This narrative is adequately captured in the following extract:

Chief’s pride at his GCHQ heroes;

Fightback on leak smears

THE boss of GCHQ has spoken for the first time of his ‘fierce pride’ in his spies after six months of smear attacks on them.

The under-fire eavesdropping agency’s director Sir Iain Lobban mounted a passionate fightback against ex-CIA technician Edward Snowden’s leaks.

Sir Iain said his brave agents had ‘definitely saved the lives’ of British troops in Iraq, Afghanistan and Libya.

And he revealed GCHQ listeners and code-breakers have won 223 campaign medals fighting alongside soldiers.

Sir Iain insisted: ‘They spend their lives protecting the security of Britain and the safety of British citizens’. (‘Chief’s Pride at His GCHQ Heroes’, *The Sun*, 7 November 2013)

The description of intelligence agents as ‘GCHQ heroes’ offers a positive representation that conforms to a pattern within tabloid discourse in which newspapers ‘swear allegiance to the virtues of patriotism and are unequivocal about the bravery of British soldiers abroad’ (Conboy, 2006: 50). This is particularly evident in the reference to campaign medals, which implies the agents possess integrity through their association with soldiering. The connection to national security is clear and strengthened by the reporting

verb 'insisted', which adds sincerity and credibility to the quotation from the director of GCHQ.

A final reporting strategy related to the defence of the rights for surveillance is to downplay the significance of Snowden's revelations. *Fuss* implies that concern about surveillance is unnecessary, for example: 'This Edward Snowden joker. Why all the fuss?' *Dreary* refers to the process of monitoring email correspondence; for example, 'Hope spooks enjoy my dreary emails' suggests that readers' emails are unlikely to be of interest to spies, and therefore, they are unaffected by the story. This is a clear contrast to the style of *The Guardian*, which elaborates on the extent and severity of the surveillance and its far-reaching consequences.

Reisigl and Wodak (2001: 44–46) outline how the way that groups and individuals are named influences how they are perceived. They explain that these 'referential strategies' place those named within social categories and emphasise certain identities over others. Once these 'social actors' are allocated to a category, this category may then be assigned linguistic 'predications': descriptions that attribute certain qualities and traits. Evidence of these strategies can be found within *The Sun* corpus keywords; for example, the words most frequently used to describe Edward Snowden include *leaker* (n=22), *traitor* (n=17), *geek* (n=7), *rebel* (n=4) and *dropout* (n=2), all of which function as referential strategies with various effects, such as to place him within a social category (*geek*) or label him a term which assumes guilt and intent (*traitor*). The girlfriend of Edward Snowden is mentioned frequently by *The Sun*, which suggests that it shares similar news values to the *Daily Mail* in terms of personalisation and sexuality. These terms relate to her physical appearance: *sexy*, *seminaked*, or define her in terms of these traits: *lover*, *beauty*, which shows how aspects of an individual's identity can be foregrounded or backgrounded through naming strategies.

The use of the keyword *quiz* to describe the 9 hour detention of David Miranda, the partner of the journalist initially contacted by Edward Snowden, is an effective example of how naming can shape representation. For example, the headline of an article reads 'May OKs cop quiz', in reference to the Home Secretary Theresa May expressing her awareness of the detention. The casual language suggests such a 'cop quiz' is nothing unusual and requires a simple 'OK' in order to be carried out. This presents his detainment as a set of quick questions asked with no ulterior motive and contrasts radically with the collocations of 'detention', the legal term for the event and lexical choice of *The Guardian*. *The Sun* refers to the release of documents as *leaks*, a word that has a clearly negative semantic prosody in the BNC and has collocations with dangerous substances such as 'nuclear', 'radioactive' and 'toxic'. This contrasts with *The Guardian* keyword for the same event: *revelations*, which has religious connotations of 'faith', 'divine', 'god' and 'truth' to form a more positive semantic prosody.

Another naming strategy of *The Sun* is its description of enemies of the state as *fanatics*, *maniacs*, *jihadis* and *plotters*. An expanded concordance of *maniacs* is shown in the following extract:

Guardian treason helping terrorists

THERE are thousands of deranged Islamist terrorists living in Britain, waiting for a chance to kill and maim the rest of us.

Semi-literate bearded jihadi maniacs piling up the ingredients in their bedsits to make bombs. ('*Guardian* Treason Helping Terrorists', *The Sun*, 10 October 2013)

The passage employs hyperbole to exaggerate the threat of terrorists, whose stated motivation is to kill 'us': the British people and implied audience of the text. Predicative naming strategies 'deranged' and 'semi-literate bearded' activate schema of how a stereotypical terrorist is imagined and attribute negative traits to this representation. The contrast between innocent British citizens and evil, murderous terrorists creates a division between the two groups into 'us' and 'them', an ideologically significant distinction.

Discussion

In addressing our first research question, keyword analysis has revealed both similarities and differences in the way the story is presented by the newspapers. *The Guardian* focused on privacy issues and the surveillance itself as the main topics and criticism of surveillance and justification of reporting as the main stances towards the story. The *Daily Mail* contained a lot of neutral keyword, and featured a number of articles outlining the surveillance, particularly on American citizens. However, the *Daily Mail* also focused heavily on Edward Snowden as an individual, including aspects of his exotic lifestyle, personal life and personality. *The Sun* was clearly opposed to *The Guardian* in its stance, being highly defensive of the rights for surveillance and critical of *The Guardian*'s reporting. It was also similar to the *Daily Mail* in the sense that it focused on the girlfriend of Edward Snowden as newsworthy.

We now return to our second research question relating to newspaper ideology. It is evident that the same original story from the same initial source has been retold and presented in substantially different ways. A reader with no prior knowledge of the events could get a very different perspective on the story depending on which newspaper they followed. To explain these differences, it is necessary to consider the function of journalism itself. Richardson (2007) highlights the importance of journalism as a means of expressing the views of the powerful, generating profit and providing entertainment, but suggests the primary function of journalism is to help readers 'better understand their lives and position(s) in the world' (p. 7). This viewpoint indicates that if readers use news to negotiate their position in the world and understand their lives, they are likely to identify with, or react to, the varying attitudes, opinions and stances taken by journalists within different news organisations and relate more strongly to certain forms of journalistic discourse. It suggests that while factors such as the social class and demographic profile of readers have some impact on the way news is structured, an attempt to identify with a more abstract, implied audience is also important. This is to some extent reflected in the findings of the corpus analysis that show evidence of relatively consistent stances and reporting strategies. However, to interpret *how* these keywords indicate news values and subtly reinforce a set of shared beliefs, a social cognition, it was necessary to contextualise them within the pre-existing knowledge frames, or context models (Van Dijk, 2008), of journalists writing for certain types of audience and analysts who are interpreting these texts. To explain the keywords, it was

necessary to draw on our own pre-existing knowledge frames for these newspapers and their position within the wider discourse of the British media. It is only with reference to such context models that we were able to provide an account of the rhetorical purposes fulfilled by the choice of particular keywords that we have described as 'reporting strategies' and demonstrate the work of social construction engaged in by journalists. The analysis of such strategies is therefore as much dependent upon these context models as it is on the identification of the keywords themselves.

The Guardian, for example, was responsible for the initial disclosures and in order to sustain its journalistic integrity needed both to justify the publication of classified documents and to prepare a defence against likely attacks. The reporting strategy employed is therefore to be highly critical of surveillance and illustrate how it directly affects its readership, with an intended effect of creating a shared set of beliefs towards surveillance between the paper and its readership. This can be seen as ideological, since the way the story is framed and structured perpetuates this shared set of beliefs. Ideologies are not inherently negative and depend on the point of view of the individual and situations they are interpreted within (Van Dijk, 2001: 12). However, such a stance is keeping with the readership's mental model for a newspaper with liberal values.

The consistent personalisation within the *Daily Mail* can also be seen as a form of news construction in which the perceived preferences of the imagined readership, their mental models, are catered for. Sparks and Tulloch (2000) demonstrate how personalisation 'can divert attention from serious issues' through focus on individuals and 'voyeuristic thrills', which suggests the strategy of focusing on Snowden's personal life may have the effect of obscuring issues of privacy and civil rights by drawing on the idiosyncratic mental models of the readership (p. 260). This is particularly evident within articles that discuss aspects of his personality, such as the 'girlish figure' article, as well as other articles within the corpus that include an evaluation of his handwriting and discussion of his attractiveness. Another interesting facet of the *Daily Mail* coverage was the reluctance to report the surveillance of the British intelligence agency GCHQ. Although the surveillance of the NSA was a relatively frequent topic, GCHQ appeared as the second most significant *negative* keyword: a word used statically *infrequently* in comparison with the reference corpus. This reluctance to report on British state surveillance may be motivated by an ideological appeal to nationalism, a strategy which Conboy (2006: 51) illustrates is used by the *Daily Mail* on other occasions to proclaim loyalty towards the armed forces.

The Sun also promotes a nationalist ideology, with a vocal defence of surveillance, strong criticism of *The Guardian* and the use of naming strategies to portray the participants of the story in a way that positions them within social groups or attributes them with certain characteristics. The division of 'us' and 'them', evident in the naming and emotionally rousing language, is the clearest example of ideology as it reinforces a mental model for nationalism. Leudar et al. (2004) examine the representation of major terrorist attacks by politicians and note how the formation of 'us' and 'them' categories result in 'the making of an enemy' (pp. 245–251). The lexical choices of *The Sun* imply that surveillance is a necessary tool for protecting the nation against terrorism. This set of beliefs contains the implicit suggestion that questioning surveillance is unpatriotic and potentially damaging to the integrity of the country. Fairclough (2010) comments that

language relating to terrorism is often vague, which makes ‘construal of an enemy’ deliberately broad and open to interpretation (pp. 488–490). Although this analysis is of a security document rather than a newspaper, the comparisons are clear and demonstrate how the vague but emotionally powerful reporting strategies of *The Sun* propagate an ‘us’/‘them’ ideology which seeks to provide justification for state surveillance and the potential restriction of civil liberties.

Conclusion

In conclusion, it is evident that each newspaper shows evidence of ideology and reports the same events in markedly different ways. However, Fowler (1991) notes that some form of perspective is present in *all* forms of discourse: He explains language is not an entirely neutral medium and that to some extent ‘values are in the language already’, making an entirely objective report difficult to achieve (pp. 41–42). When writing this article we used the term ‘surveillance’ in an attempt to describe the act of monitoring the public neutrally, yet ‘surveillance’ may activate an Orwellian mental model. Many other words could have been used, such as ‘snooping’, ‘spying’, ‘eavesdropping’, ‘listening’ and ‘investigating’ – all of which have various connotations to different readers depending on their mental models. This shows that it is difficult to avoid conveying an ideology and stance in writing, due to the inherent, and sometimes personal, associations embedded within words.

Despite this, the fact that ideology exists so clearly in journalistic discourse is significant: It creates substantially different representations of events and highlights or obscures various attitudes, issues and values. Richardson (2007) writes that journalistic language is ‘*intimately* related to power’, and this is important because ‘journalism has more power to shape our understanding about events, ideas, people and the relationships between people, than many other forms of communication’ (p. 220). However, it is important to note that readers do not play an entirely inactive role in understanding texts. Readers bring different mental models to their reading of a text and do not simply digest the viewpoints within news but react to them in various ways (Richardson, 2007: 41). Fowler (1991) explains how this gives more power to readers to determine their own opinions, stating that newspapers do not ‘consciously wrap’ selected events in ideological language which the reader then ‘passively absorbs, ideology and all’ (p. 41). This is a significant point because it shows that as consumers of news, we have the ability to detect overtly ideological news and critically engage with it to determine for ourselves which elements we agree with or which parts we find questionable.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. **[AQ: 4]**

Notes

1. A possibility we considered was to compare each research corpus against the discourse reference corpus with that particular research corpus excluded; for example, *The Sun* research

corpus could be compared against the discourse reference corpus with *The Sun* corpus taken out. However, a pilot study using a 100,000 word corpus of speeches of Tony Blair compared with a 400,000 word corpus of British speeches *excluding* those by Blair's (Method A) and a 500,000 word corpus of British speeches *including* those by Blair (Method B) showed that 96 of the top 100 keywords identified in Method A were also among the 500 top keywords using Method B; so whichever method is used the keywords are largely similar although their order is slightly different. The advantage of Method A is that it can identify keywords more accurately, but the disadvantage is that it requires changing the discourse reference corpus for each comparison that is made. For reasons of methodological simplicity, we therefore decided to use Method B as it keeps the discourse reference corpus constant in line with its role as a control corpus.

2. Keywords are shown in italics throughout.

References

- Anthony L (2011) *AntConc*, version 3.2.2 (Computer Software). Tokyo, Japan: Waseda University. Available at: <http://www.antlab.sci.waseda.ac.jp/> (accessed January 2014).
- Baker P (2006) *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker P (2014) *Using Corpora to Analyze Gender*. London; New York: Bloomsbury Academic.
- Bednarek M and Caple H (2014) Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond. *Discourse & Society* 25(2): 135–158.
- Caple H (2013) *Photojournalism: A Social Semiotic Approach*. Basingstoke: Palgrave Macmillan.
- Charteris-Black J (2012) Comparative keyword analysis and leadership communication: Tony Blair – A study of rhetorical style. In: Helms L (ed.) *Comparative Political Leadership*. Basingstoke: Palgrave Macmillan, pp. 142–164.
- Charteris-Black J (2014) Political style: A study of David Cameron. In: Stockwell P and Whiteley S (eds) *The Handbook of Stylistics*. Cambridge: Cambridge University Press. **[AQ: 5]**
- Conboy M (2006) *Tabloid Britain: Constructing a Community through Language*. Oxon: Routledge.
- Cotter C (2001) Discourse and media. In: Schiffren D, Tannen D and Hamilton HE (eds) *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishing, pp. 416–436.
- Cotter C (2010) *News Talk: Investigating the Language of Journalism*. Cambridge: Cambridge University Press.
- DeWerth-Pallmeyer D (1997) *The Audience in the News*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fairclough N (2010) *Critical Discourse Analysis: The Critical Study of Language*, 2nd edn. Harlow: Pearson Education.
- Fowler R (1991) *Language in the News: Discourse and Ideology in the Press*. London: Routledge.
- Harcup T and O'Neill D (2001) What is news? Galtung and Ruge revisited. *Journalism Studies* 2(2): 261–279.
- Hunter D and MacDonald MN (2013) The discourse of Olympic security: London 2012. *Discourse & Society* 24(1): 66–88.
- Kim KH (2014) Examining US news media discourses about North Korea: A corpus-based critical discourse analysis. *Discourse & Society* 25(2): 221–244.
- Leudar I, Nekvapil J and Marsland V (2004) On membership categorization: 'Us', 'Them' and 'Doing Violence' in political discourse. *Discourse & Society* 15(2–3): 243–266.
- LexisNexis (2014) Available at: <http://www.lexisnexis.com/uk/nexis/> (accessed January 2014).

- Purvis T and Hunt A (1993) Discourse, ideology, discourse, ideology, discourse, ideology. *The British Journal of Sociology* 44(3): 473–499.
- Reah D (2002) *The Language of Newspapers*, 2nd edn. London: Routledge.
- Reisigl M and Wodak R (2001) *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. London: Routledge.
- Richardson JE (2007) *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Basingstoke: Palgrave Macmillan.
- Seale C and Charteris-Black J (2010) Keyword analysis: A new tool for qualitative research. In: Bourgeault IL, DeVries R and Dingwall R (eds) *The SAGE Handbook of Qualitative Methods in Health Research*. London: SAGE, pp. 536–556.
- Sparks C and Tulloch J (2000) *Tabloid Tales: Global Debates Over Media Standards*. Lanham, MD: Rowman & Littlefield.
- Sumpter R (2000) Daily newspaper editors' audience construction routines: A case study. *Critical Studies in Media Communication* 17(3): 334–346.
- Van Dijk T (2001) Discourse, ideology and context. *Folia Linguistica* 35(1–2): 11–40.
- Van Dijk T (2008) *Discourse and Context: A Sociolinguistic Approach*. Cambridge: Cambridge University Press.

Author biographies

Jens Branum is a graduate of English Language and Linguistics from the University of the West of England.

Jonathan Charteris-Black is Professor of Linguistics at the University of the West of England, UK. He is the author of *Corpus Approaches to Critical Metaphor Analysis* (Palgrave, 2004), *Politicians and Rhetoric* (Palgrave, 2005 and 2011), *The Communication of Leadership* (Routledge, 2007), *Gender and the Language of Illness* (Palgrave, 2010), *Analysing Political Speeches* (Palgrave, 2014), as well as numerous other articles and book chapters.