

Investigating evaluation and news values in news items that are shared through social media

Monika Bednarek¹

Abstract

The sharing of news through social media platforms is now a significant part of mainstream online media use and is an increasingly important consideration in journalism practice and production. This paper analyses the linguistic characteristics of online news sharing on Facebook, with a focus on evaluation and news values in a corpus of the 100 ‘most shared’ news items from ‘heritage’ English-language news media organisations. Analyses combine corpus linguistic techniques (semantic tagging, frequency analysis, concordancing) with manual, computer-aided annotation. The main focus is on discursive news values analysis (DNVA), which examines how news values are established through semiotic resources, enabling new empirical insights into shared news and adding a specific linguistic focus to the emerging literature on news sharing. Results suggest that all ‘traditional’ news values appear to be construed in the shared news corpus and that there is variety in terms of the items that are widely shared. At the same time, the news values of Eliteness, Superlativeness, Unexpectedness, Negativity and Timeliness seem especially important in the corpus. The findings also indicate that ‘unexpected’ and ‘affective’ news items may be shared more, and that Negativity is a more important news value than Positivity.

Keywords: Corpus Linguistics, discursive news values analysis, evaluation, news sharing, news values, news discourse, newsworthiness, online audience metrics, online social media, shared news.

¹ Department of Linguistics, University of Sydney, John Woolley Building A20, NSW 2006, Australia.

Correspondence to: Monika Bednarek, *e-mail:* monika.bednarek@sydney.edu.au

1. Evaluation and news values

Thompson and Hunston (2000: 6–7) argue that one of the three functions of evaluation is to reflect a value system: ‘Every act of evaluation expresses a communal value-system, and every act of evaluation goes towards building up that value-system. This value-system in turn is a component of the ideology which lies behind every text’. Importantly, as they point out, this value system or ideology may belong to a sub-group, including members of a particular profession such as ‘applied linguists’. The value system that I am exploring in this paper is that at play in the professional context of journalism. In much of the relevant literature (see Caple and Bednarek, 2013) it is assumed that journalists measure and judge the perceived newsworthiness of events according to a set of news values such as Negativity, Unexpectedness and Proximity. News values hence make up a professional value system which reflects dominant societal ideologies (Bell, 1991: 156). These news values are also discursively constructed in published news stories (Bednarek and Caple, 2012a). This is because it is arguably a central goal of news stories to present events as ‘newsworthy’ – that is, as conforming to the news values. While the concept of news values originates from outside linguistics (Galtung and Ruge, 1965), it is, therefore, an important consideration for any linguistic analysis of news texts. Discursive news values analysis (DNVA), which is further explained below, can be used by linguists to analyse how news values are constructed through discourse in news texts – for example, through evaluative language (Bednarek, 2006) or through what Bell (1991: 177–80) has called the ‘value-laden’ lexicon of newsworthiness.²

The particular types of news texts that I am considering in this paper are shared online news texts: news items that are published on the web and whose content and internet location (URL) are shared with others by users of social media networks. Some have argued that a new evaluative regime is emerging in the age of online news sharing (Crawford *et al.*, 2015). This paper takes a first step in the linguistic analysis of news values in shared news online, employing DNVA to systematically examine news texts for constructed news values, and hence enabling new empirical insights into shared news.

While it is beyond the scope of this article to give a detailed outline of DNVA³ (see Bednarek and Caple 2012a,b, 2014, forthcoming; Caple, 2013; and Potts *et al.*, 2015) I list three basic tenets here:

- News values are defined as the ‘newsworthy’ aspects of events in line with the news values of Consonance, Eliteness, Impact,

² The ‘discursive perspective’ (Bednarek and Caple, 2012a) on news values differs somewhat from the perspective taken in Journalism Studies, where news values are usually regarded as a production technique that news organisations use to systematise and determine what will be researched or reported.

³ See <http://www.newsvaluesanalysis.com>.

Negativity, Personalisation, Proximity, Superlativeness, Timeliness and Unexpectedness (Table 1). This list is based on a review of the literature on news values (Cagle and Bednarek, 2013).

- These news values are linked systematically to verbal and visual semiotic resources. Table 1 provides an updated summary of verbal resources, with more complete versions to be published elsewhere (Bednarek and Cagle, forthcoming).
- This provides a framework for analysis—using corpus linguistic methods, for instance, it becomes possible to identify from frequency lists those forms that may potentially establish newsworthiness. Such forms are labelled ‘pointers’ to newsworthiness (Bednarek and Cagle, 2014: 145). However, this framework (and Table 1) cannot be taken as an automatic checklist. Rather, close attention needs to be paid to the meaning potential of the linguistic resource as used in a news story, as well as to the target audience and time/place of publication.

While Positivity is not included in Table 1, because it is only recognised by some scholars (see Cagle and Bednarek, 2013: 19) and may only play a role in some types of news (Bednarek, 2016), this paper will consider both Negativity and Positivity in order to explore whether news constructed as negative or positive is shared more by users. Positivity can be constructed through similar resources as Negativity, albeit with the opposite valence.

As Table 1 indicates, news values are not just constructed through language that we might call ‘evaluative’, although this clearly depends on how evaluation and evaluative language are defined. Thompson and Hunston (2000: 21) propose that evaluation can be identified in texts by looking for comparators (including comparative adjectives/adverbs, adverbs of degree, comparator adverbs and expressions of negativity), markers of subjectivity (e.g., modals) and markers of value (e.g., evaluative lexis, adjectives, adverbs, nouns, verbs and indicators of the existence of goal achievement). In this sense, some of the devices listed in Table 1 are examples of evaluative language, including evaluative lexis, quantifiers, intensifiers and comparison. Other devices—such as place and time references or tense and aspect—would not normally be classified as evaluative language (but see Hood, 2010), although they can be regarded as ‘markers of value’ in a broader sense, in that they are indications of the construction of a professional value system (news values), as explained above.

Understanding the construction of news values in shared online news is important because social media services are becoming increasingly significant for the production, dissemination and consumption of online news (Olmstead *et al.*, 2011; Martin and Dwyer, 2015, and Weeks and Holbert, 2013). Social media analytics, which measure the time, location, audience and nature of news consumption, are having a huge impact on news

News value	The event is constructed as...	Linguistic resources
Consonance	stereotypical	constructions of stereotypes, for example through evaluations of expectedness (<i>notorious, famed for...</i>) or similarity with past (<i>typical style, once again...</i>)
Eliteness	of high status or fame (including news actors, organisations, etc.)	various status markers, including labels, recognised names, evaluations of importance, descriptions of achievement (<i>experts at Harvard university, a high-profile arrest, Barack Obama, the Oscars, a key federal government minister, top diplomats, were selling millions of records a year...</i>)
Impact	having significant effects or consequences (not limited to impact on the target audience)	evaluations of significance (<i>momentous, historic, crucial...</i>); reference to real or hypothetical important or relevant consequences (<i>note that will stun the world, Australia could be left with no policy, leaving scenes of destruction...</i>)
Negativity	negative	negative evaluative language (<i>terrible, dangerous, slaughter...</i>); reference to negative emotion and attitude (<i>distranght, worried, condemn, criticise...</i>); negative lexis (<i>conflict, damage, death, crisis, abuse, controversial, row...</i>), other references to negative happenings (e.g. the breaching of socially approved behaviour/norms)
Person-alisation	having a personal or 'human' face (involving non-elite actors)	references to 'ordinary' people, their emotions, experiences (<i>Charissa Benjamin and her Serbian husband, 'It was pretty bloody scary'; But one of his victims [...] sobbed, Deborah ... said afterwards: 'My sentence has only just begun'...</i>);
Proximity	geographically or culturally near	explicit references to place or nationality near the target audience (<i>an Australian; Australia, Canberra woman...</i>); references to the nation/community (<i>the nation's capital; home-grown...</i>); inclusive first-person plural pronouns (<i>our nation's leaders...</i>); cultural references (e.g., <i>prom, haka</i>)
Superlativeness	of high intensity or large scope/scale	quantifiers (<i>many, all, thousands...</i>); intensifiers (<i>sensational, dramatically, super, severe, extreme...</i>), including intensified lexis (<i>epidemic, smashed, stun, wreck...</i>); references to growth/escalation (<i>growing, raised...</i>), repetition (<i>building after building...</i>); some instances of metaphor/personification/simile (<i>a tsunami of crime; epidemic swallowing Sydney, looked like the apocalypse...</i>); comparison (<i>biggest counter-terrorism raid, most shocking child abuse case...</i>).
Timeliness	timely in relation to the publication date: new, recent, ongoing, about to happen, current or seasonal	indications of newness or change (<i>fresh, new, latest, for the first time...</i>); explicit time references (<i>today, yesterday, within days, now...</i>); implicit time references (<i>continues, ongoing, have begun to...</i>), selected verb tense and aspect (<i>have been trying, is preparing, is about to...</i>); references to seasonal or current happenings/trends
Unexpectedness	unexpected	evaluations of unexpectedness (<i>different, astonishing, strange...</i>), comparisons that indicate unusuality (<i>the first time since 1958, Sydney's wettest August in 16 years</i>); references to surprise (<i>shock at North Cottesloe quiz night; people just really can't believe it...</i>); references to unusual happenings (<i>British man survives 15-storey plummet...</i>).

Table 1: Linguistic resources for construing newsworthiness.

production and have transformed twenty-first century newsrooms alongside other technological changes:

Today, the ‘new newsroom’ has optimisation desks, to make stories work better on social media, data scientists who analyse the information about story performance to tell journalists how to write headlines, produce photographs and report stories which will be ‘liked’ and ‘shared’ more than others. It has aggregation desks, which scour the web to find news that ordinary people have posted for a wider audience. It has audience insight desks that work on how to get more people to spend longer reading more journalism. And it has data desks, which take the newly available sources of information in vast quantities and use the latest mining tools and techniques to clean, interpret and visualise information in new ways.

(Bell, 2015)

Surprisingly, given their centrality to contemporary journalism, there is still little analysis of news values and social media networks (Facebook and Twitter), in particular in relation to shared news. In his study of news framing on Twitter, Wasike (2013: 11, 20) claims that ‘no studies have looked specifically at [...] news values vis-à-vis media format as pertaining to social media’.⁴

However, some non-linguistic research exists into the content of shared news. In a study of ‘most emailed’ *New York Times* articles, Berger and Milkman (2011: 5–6) coded articles into positive/negative and evoked emotions (anxiety, anger, awe and sadness) and argue that affect-laden content is more viral than non-affective content, while positive articles are more viral than negative articles. However, what they call ‘high-arousal’ emotions are more viral than others, regardless of positivity/negativity. More specifically, they claim that content that evokes positive high-arousal emotions (their examples include awe and high amusement) and negative high-arousal emotions (their examples include anger and anxiety) is more viral than content that evokes ‘more of a deactivating emotion (i.e., sadness)’ (Berger and Milkman, 2011: 10). In addition, articles that are interesting/surprising and practically useful (informative) are more likely to be ‘most emailed’. In contrast to Berger and Milkman’s study, Hansen *et al.*’s (2011: 12) study of tweets suggests that ‘negative news is more retweeted than positive news’, and they conclude that negativity promotes retweeting

⁴ A considerable amount of non-linguistic research has explored the use of social media by audiences for news consumption and sharing (e.g., Bruns *et al.*, 2013; Hermida *et al.*, 2012; and Olmstead *et al.*, 2011), especially predictors of, reasons for and causes of sharing information (such as information seeking, status seeking, socialising, entertainment and physiological arousal), often in relation to social sharing in general, and which may include the sharing of news items (e.g., Berger, 2011; Lee and Ma, 2012; and Weeks and Holbert, 2013). For a contribution from linguistics on ‘share-worthiness’ in the context of list-sites, see Pflaeging (2015).

and virality in news items, but not in non-news. To explain their differing result, Hansen *et al.* (2011) argue that there might be differences between e-mail networks and Twitter networks.

In an analysis of most shared UK news stories on Twitter (from BBC, Channel 4, *Guardian*, Sky and *Telegraph*), Newman (2011: 22–4) proposes that the items that tend to do well are breaking news and original, distinctive content. In relation to content, he identifies quirky/funny items, items relating to disasters/deaths and provocative comment/analysis. In addition, analysis of ‘mood’ suggests that most shared stories involve shock/surprise or are funny/weird. Finally, Crawford *et al.* (2015) provide an industry perspective on shareability, arguing for four qualities that make content shareable: simple, emotional, new/unexpected and triggered (prompted by something). Because of their disciplinary origin, these and similar analyses tend not to report any findings on specific linguistic features of shared news, as they are focussed on coding content, often automated through tools such as sentiment analysis, or on the significance of these changing practices for the media industry and their attendant policy debates. This paper thus adds a specific linguistic focus to the emerging literature on news sharing.

2. Corpus design

For this study, I draw on a corpus compiled by a team for a bigger project on shared news online (see Acknowledgments). As part of this project, we compiled a small corpus of news items that were shared many times among users through Facebook. Since 2005, Facebook has developed different functionalities that facilitate the sharing of content by its users, as summarised by Pflaeging (2015). We chose Facebook over other social networks such as Twitter because Facebook is ‘by far the most important network for news everywhere’ (Reuters Institute Digital News Report, 2014: 8) and has been called the ‘news powerhouse among the social media sites’ (Anderson and Caumont, 2014). More specifically, we focus on shared generalist news items that originate with print and broadcast English language ‘heritage’ news media organisations (such as *New York Times*, *Guardian*, CNN and FOX) rather than ‘digital natives’ (online-only new media organisations such as BuzzFeed, Upworthy and *Huffington Post*).⁵ The business model of the latter organisations focusses on promoting news sharing and they employ a greater array of techniques and work practices to encourage this behaviour. They also create content that would not traditionally be called ‘news’ but is widely shared, such as funny cat videos. We also excluded content from magazines such as the *Atlantic* and *Time*, and specialist publications such as *Forbes* and *Hollywood Reporter*.

⁵ We also included news.com.au, a news site bringing together journalism from News Corp Australia.

To compile the corpus, we used Share Wars' Likeable Engine analytics software⁶ to extract the top 200 items by total Facebook share count as at early September 2014. The Likeable Engine frequently (at least every twenty minutes) and automatically collects URLs from selected news media homepages (primarily based in the US, UK and Australia), tracking their Facebook share counts. Our corpus uses URLs collected from October 2011 to July 2014, although the exact set of homepages collected by the Likeable Engine has varied over this period. Since Facebook only provides a snapshot of the share count, rather than a history, we captured the count at least three months after publication, assuming that most news items stop accruing shares after a short period. Thus, we based our data collection on a 'final' Facebook share count at early September 2014, selecting the top 200 URLs by this measure that are not excluded by the criteria above.⁷ These 200 URLs are arguably representative of the types of heritage news items that are shared widely by Facebook users.

Manual examination of these URLs showed a wide variety of journalistic text types: hard news, soft news, research news, feature, biography, eulogy/obituary, explainer (including a graphic), image gallery and photo essay, quiz, test, letter, e-mail, opinion (including advice, analysis, cartoon and list), personal and research recount, and wish list. Presumably, even more text types (cat videos, *etc.*) would be identified if shared URLs from digital natives were included. In fact, the digital environment has resulted in a diversification and proliferation of journalistic text types that did not occur in print news, and that deserve further linguistic attention, both in relation to their generic structure and their linguistic characteristics. Each of these merits closer analysis, and rather than asking what words/phrases are frequent in shared items in general, it may make more sense to ask, for instance, what quizzes, what image galleries and what opinion are shared, because these genres are so different from one another. Thus, we decided to focus first on items that are traditionally regarded as news stories, excluding any visual-centric items (e.g., photo essay and image gallery), opinion and analysis (e.g., explainer and argument), interactives (e.g., quiz and test), features, obituaries, interviews, and so on. News stories were included in the corpus if they described an event, happening or issue concerning other participants and if the reported event was either new or a new development. We made our decision on the basis of visiting each of the 200 URLs and assessing each item.⁸ Note that some items excluded from our corpus may be regarded as 'news' in other contexts, depending on how news is

⁶ For further information, see <http://likeable.share-wars.com/> and Martin and Dwyer (2015).

⁷ Thus, most excluded URLs are from online-only, digital native sites (e.g., buzzfeed.com) and online versions of magazine publications (e.g., time.com). Domain root URLs (e.g., <http://cnn.com/> rather than an article within that domain) also appeared in the Likeable Engine listing but were removed. Four URLs that ranked within the top 200 could not be fetched.

⁸ We also identified two additional root URLs, which were excluded.

defined and what news genres are recognised. However, it was necessary to operationalise news items so that we could compare ‘like with like’, as it were, and to exclude any items that do not lend themselves to DNVA, as their communicative function is not to construct events as newsworthy.

It was also necessary to operationalise items using language-external criteria, as far as possible, to avoid circularity of findings (Sinclair, 2004). We therefore attempted to operationalise news items with regard to their communicative function, rather than through specific micro-linguistic features such as presence/absence of evaluative language. We also took into account in what section of the website an item occurred (e.g., World or Opinion) and how it was labelled by the website itself (e.g., ‘on parenting’) – although this categorisation was not always reliable. However, it is clear that this is not an exact science and we assessed the communicative function of texts on the basis of reading them. Thus, the selection was not entirely independent of language. After selecting items from the top 200 using this procedure, the Shared News Corpus (SNC) consists of almost 100 news items, mainly hard news (including reports of the death of famous people), soft news and research news.⁹ The decision to start with a small corpus was deliberate, because it allows us to combine quantitative and qualitative corpus and discourse analytical techniques, which can inform later analyses of larger corpora.

Despite the limitations noted above, the corpus appears representative: all items come from English-language news media organisations across four different cultures (UK, USA, Australia and New Zealand), and were successful in terms of their Facebook share count, as specified above. Table 2 lists the news organisations, countries and number of items in the corpus, and shows that the SNC is US-centric and comprises both ‘quality’ and ‘popular’ news.

3. Corpus linguistic analysis

This study combines corpus linguistic techniques (semantic tagging, frequency analysis and concordancing) with manual, computer-aided

⁹ Different suggestions have been made to distinguish hard news and soft news – for example, hard news has been classified as time-bound (Bell, 1991) or as destabilising (Feez *et al.*, 2008) in contrast to timeless (Bell, 1991) and stabilising (Feez *et al.*, 2008) soft news. In our experience, these criteria cannot always be consistently applied, but in any case we are not concerned with exploring differences between hard and soft news. News stories that primarily deal with telling how someone died were included, but those items that were primarily about the person who died were not included, as they did not report on the event but, rather, described the person’s life (obituary). Most of the death reports in the corpus do have some description of the person’s life but also include a significant amount of reporting on the actual event leading to the death. Further, research news was included (journalists writing about the findings of an academic study), but not research recounts (researchers writing about their own research, in the first person).

News organisation and country	No. of items in corpus
CNN (USA)	23
FOX (USA)	20
<i>The Daily Mail</i> (UK)	13
<i>The New York Times</i> (USA)	7
<i>Washington Post</i> (USA)	6
<i>USA Today</i> (USA)	5
<i>The Guardian</i> (UK)	4
<i>Washington Times</i> (USA)	4
<i>The Daily Telegraph</i> (UK)	3
ABC (Australia)	2
BBC (UK)	2
News Corp (Australia)	2
NBC (USA)	1
<i>The Daily News</i> (USA)	1
<i>The Independent</i> (UK)	1
<i>Miami Herald</i> (USA)	1
<i>The Mirror</i> (UK)	1
<i>The Age</i> (Australia)	1
<i>San Francisco Gate/Chronicle</i> (USA)	1
<i>New Zealand Herald</i> (NZ)	1
Total	99

Table 2: News organisations.

annotation. More specifically, the corpus was tagged using the UCREL Semantic Analysis System (USAS; Archer *et al.*, 2002). Each semantic tag denotes the correspondence of an item to one of twenty-one discourse fields, such as ‘Emotional Actions, States and Processes’ or ‘Time’ with 232 subdivisions such as ‘Happy/sad: Happy’ or ‘Time: General: Future’. In other words, the tagging is based on ‘semantic fields which group together word senses that are related by virtue of their being connected at some level of generality’ (Archer *et al.*, 2002: 1). For example, the items *recent*, *latest* and *new* all belong to the semantic field labelled ‘Time: Old, new and young; age’

and are tagged as such. Using a corpus of American newspaper texts, Potts (2013) found that in approximately 85 percent of cases the semantic tag listed first in the string of candidates and deemed ‘most likely’ by the USAS tagger was appropriate. I worked with a list of the most frequent semantic tags, based on the first sense tag and the first domain when multiple tags/domains were assigned and treating multi-word expressions as single tokens. Analysis was restricted to the top 100 most frequent tags.

Further, various frequency lists were extracted from the corpus: a list of word forms, lemmas and *n*-grams (bigrams, trigrams and tetragrams).¹⁰ The lists of word forms and *n*-grams were sorted according to ‘range’ (the number of corpus files that a word form occurs in), and the lemma list was sorted according to frequency. For reasons of scope, analysis of word forms, lemmas and *n*-grams was then restricted to those occurring across at least twenty (of ninety-nine) items.¹¹ This limit is important because I am interested in occurrences across a range of shared news texts, rather than those occurring in only one or a few items. Applying such a limit allows insights into the linguistic characteristics of shared news more generally.¹² However, this means that very few *n*-grams are included in the analysis, since most do not occur across at least twenty texts.

Following a similar procedure as that suggested in Bednarek and Caple (2014) and Potts *et al.* (2015), all lists (word forms, lemmas, *n*-grams and semantic tags) were examined to identify items and tags that might be ‘pointers’ to a specific news value, based on their meaning potential (but not their actual use, as comprehensive concordancing was only undertaken if explicitly mentioned below). To identify pointers, I consulted the linguistic framework for DNVA introduced above as well as the USAS manual and the prototypical examples for each tag provided in the manual (Archer *et al.*, 2002). In discussing the results, I will incorporate some reflections on the usefulness of these corpus techniques for DNVA.

For reasons of scope, I excluded most grammatical items such as *the*, *a*, *to* and *of*, and grammatical tags such as Z8 (Pronouns), Z6 (Negative)

¹⁰ Given that the corpus contains British and American texts, Wordsmith’s search function was used to see how many word forms were affected by spelling variation, which indicated that standardisation of spelling would not have resulted in radically different results.

¹¹ The lemma list was extracted using Yasumasa Someya’s list (40,569 tokens in 14,762 lemma groups), available at http://www.lexically.net/downloads/BNC_wordlists/e_lemma.txt. It was sorted according to frequency, since Wordsmith was not able to sort lemmas correctly according to range. Analysis of range for lemmas (as a whole, not for individual word forms) was, thus, restricted to the top 250 most frequent – occurring at least thirty-two times. Some manual correction and disambiguation of the lemma list was undertaken and taken into account in the range analysis. For example, 18 of 119 occurrences of the lemma USE were noun, rather than verb occurrences. The frequency for the lemma USE (as verb) was thus changed to 101, but the lemma USE (as a verb) was retained since it occurs at least thirty-two times and across more than twenty corpus files. In other cases, disambiguation meant that lemmas were excluded.

¹² Note that the semantic tag analysis does not take into account range.

or Z7 (*if*), because of their multi-functionality. However, I did include intensifiers/quantifiers and tags for time, since these are potentially crucial resources for Superlativeness and Timeliness. To clarify, an item identified as ‘pointer’ can be used to construct a specific news value, although it does not necessarily do so in all its uses/meanings. However, not all cases are clear-cut: at what point should an item or tag be categorised as ‘pointer’, when it has different meanings and uses? In my categorisation I followed a relatively conservative approach and in doubtful instances I did not include the item or tag as a potential ‘pointer’.¹³

I acknowledge, here, that other researchers might identify a different list of ‘pointers’ and that researcher subjectivity plays a role in DNVA, regardless of whether corpus or other techniques are used. This is also the case for the manual, computer-aided annotation which complements the corpus linguistic techniques. Here, for each of the ninety-nine corpus texts, the headline (H) and the opening paragraph (OP) were read and then coded for each news value. Together, the H and OP often act as ‘summary’ (van Dijk, 1988), ‘abstract’ (Bell, 1991) or ‘nucleus’ (Iedema *et al.*, 1994) where the reported event is represented as newsworthy (White, 1997: 128). Mahlberg and O’Donnell’s (2008) corpus study of first sentences in news stories from *The Guardian* identified patterns that interpret the topic as newsworthy. In general, the OP of a traditional news item comprises the ‘most important news element of the story in addition to the choice of angle or “hook”, or approach to the subject’ (Cotter, 2010: 162). In other words, we can expect the OP of a news story to be the structural element where we find the news values that are emphasised in the story – the ‘news values angle’. To code Hs and OPs I used the UAM Corpus Tool (O’Donnell, 2015), a software program that can be used to annotate text manually, with the software then providing the researcher with quantitative patterns. Since researcher subjectivity may be problematic here, I have, for the main part, used three coding choices for OPs and Hs: ‘yes’ (the news value is constructed), ‘no’ (the news value is not constructed), ‘possible’ (for debatable, unclear, uncertain or special cases), and have made the coding manual available (Bednarek, 2015). In reporting results for each news value below, I will first comment on findings from the identification of pointers in the frequency and tag lists, before providing results from this manual analysis of OPs/Hs.

¹³ That is, I did not classify as pointers tags such as A9 (Getting and giving: possession), X2.1 (Thought, belief), B3 (medicines and medical treatment), A13.4 (Degree: Approximators), or A6.1– (Comparative terms denoting difference) because their general potential to construct news values seemed doubtful. Further, lemmas/word forms such as SAME, TIME, HOUR, DAY, *night*, MONDAY, EARLY, CASE, LIFE, LIKE, TRY, STOP, FEEL, PUBLIC and STUDY (N) were excluded because of their high level of multifunctionality, although one might imagine contexts in which they construct a specific news value.

4. Results

4.1 Consonance

No potential pointers to Consonance can be identified from the frequency lists (word forms, lemmas and *n*-grams), while the semantic tag A6.2+ (Comparing: Usual) may indicate instances where Consonance is established in the corpus (including *common, commonly, trend, usual, natural, typical, regular, routinely, standard, normal* and *average*), although concordancing would need to confirm this. In general, it may well be the case that it is rarely individual word forms that are used to establish this news value, but rather that Consonance is construed more implicitly. This hypothesis is confirmed by the results of a previous large-scale corpus study of news discourse that did not identify clear pointers to this news value in the top 200 most frequent lemmas (Potts *et al.*, 2015: 170).

The manual, computer-aided annotation of Hs/OPs (based on reading each headline and opening paragraph and using three coding choices, as explained above) suggests that twenty headlines and twenty-seven OPs ‘possibly’ construct Consonance. OPs may provide more information than the headlines – for instance, a news actor’s national or regional origin, as in Example 1, from CNN.

- (I) Sheriff: Father kills man sexually abusing his daughter (H)
 A **Texas** father caught a man sexually assaulting his 4-year-old daughter and punched him in the head repeatedly, killing him, authorities said. (OP)

In this case, the H does not indicate where the event took place, whereas the OP identifies the location as Texas, which may conform to a potential stereotype that some members of the target audience hold about Texans – that they like to take the law in their own hands.

4.2 Eliteness

Several items from the frequency lists clearly point to Eliteness, including *official, university, government, president* and *US/United States, state* (noun), *American* (the United States as elite country/nationality),¹⁴ while the lemmas CITY, AREA, COUNTRY and GROUP would need further investigation as to whether they refer to an elite place (through endophoric or exophoric reference) or group (e.g., as part of a high-status role label such as ‘business groups’). Promising semantic tag candidates that would deserve further qualitative investigation are listed in Table 3.

¹⁴ Probably the result of the US-centric nature of the corpus.

Semantic tag and paraphrase	Some instances from the corpus
G1.1 tags (Government, etc.)	<i>authorities, parliament, congressional, officially, federally, governmental...</i>
S7.1+ (Terms depicting power/authority/influence and organisation/administration)	<i>leading, lead, managing, leadership, chief, forced, ruled, won...</i>
G2.1 (Law and order)	<i>rules, police, law, court, regulations, legal, judge...</i>
Z2 (Geographical names)	<i>Washington, Salzburg, Afghanistan, American, Irish, Israeli...</i>
M7 (Places)	<i>place, country, town, area, region, district, province, hometown...</i>
S5 tags (Groups and affiliation)	<i>crowd, group, member, team, network, mob, federal, independent, affiliate...</i>
Z3c (Other proper names)	<i>Starbucks, Pepsi, UN, Yale, RSPCA...</i>
Z1 tags (Personal names)	<i>Jim, George, Amy, Kathryn...</i>

Table 3: Semantic tag candidates for Eliteness.

Even without concordancing, however, it is clear that not all instances construct Eliteness—for example, *Afghanistan* would not be considered an elite place by the newspapers that are included in the corpus and personal names such as *Kathryn* may refer to ordinary people rather than elites. This illustrates that while semantic tag analysis does carry some advantages for DNVA (see Potts *et al.*, 2015), the disadvantage is that the semantic field is often too broad. For example, the semantic field S7.1+ comprises items relevant to the establishment of Eliteness (*leading* and *chief*), as well as items which are less likely to construe this news value (*won*). There are also differences in the status of law and order authorities (e.g., *judge* versus *police*) which are disguised by the relevant tag (S7.1+). Thus, it is not currently possible to use a semantic tagger as a news values tagger. At the same time, it is too time-consuming to examine each instance of each tagged item qualitatively, and the approach taken in Potts *et al.* (2015)—to focus on the most frequent constituents and random thinning to fifty instances—still only provides a partial picture. In this paper, the decision was taken not to follow this approach but, rather, to undertake manual analysis of Hs/OPs as a different qualitative method.

The manual analysis of Hs/OPs suggests that forty-four headlines and forty-nine OPs construct Eliteness, while twenty Hs and sixteen OPs ‘possibly’ do so. Where the headline constructs Eliteness, but the OP does not, the OP tends not to be a typical summary ‘hard news’ lead, for example:

- Emily Kraus was psyched.
- It’s an image that tugs at the heartstrings. [followed by description of image]
- Let’s get ready to . . . mumble.
- An e. You can write it with one fluid swoop of a pen or one tap of the keyboard. The most commonly used letter in the English dictionary. Simple, right?
- Many animal lovers find it hard to part with their pets when they die.
- Are you prepared for the impending zombie invasion?
- It was a scene as creepy as a Hannibal Lecter movie.

4.3 Impact

Pointers to Impact that can be identified from the frequency lists consist of the noun *result* and the verbs *leave*, *cause*, *become*, *follow* and *help*. While *help* may point to positive impact (if something/someone has helped to achieve a positive outcome), the other lemmas seem to be lexical markers of causality which establish Impact only if they construct the news event as having significant effects or consequences. With all of the Impact pointers, the co-text would, therefore, need to be examined. This is also true of the related semantic tag candidates A2.2 (Affect: Cause/Connected; e.g., *result*, *cause*, *factor*, *causal*, *link*, *induced*, *responsible* and *consequences*), A2.1+ (General/abstract terms denoting [propensity for] change; e.g., *transforming*, *became*, *occurrence*, *changed*, *make a difference*, *affected* and *happened*) and S8+ (Helping; e.g., *rescue*, *supporting*, *welfare*, *blessing*, *promoting*, *helps*, *benefit* and *aid*). However, A11.1+ (Abstract terms denoting importance/significance) may be a more promising candidate, as words tagged as such do seem to point to significance, including: *major*, *serious*, *significant*, *priority*, *fundamental*, *important*, *emergency*, *crucially*, *decisive* and *central*.

The manual analysis of Hs/OPs suggests that Impact tends not to be explicitly constructed in either headlines or OPs, with only twelve Hs and eighteen OPs doing so (plus another ‘possible’ one H and three OPs). Sometimes the OP makes the impact explicit where the headline does not, as in Examples 2 to 4.

- (2) Arizona lawmakers pass controversial anti-gay bill (H)
Arizona’s Legislature has passed a controversial bill **that would allow business owners, as long as they assert their religious beliefs, to deny service to gay and lesbian customers.** (OP)

- (3) Federal judge rules DC ban on gun carry rights unconstitutional (H)
A federal judge in the District of Columbia on Saturday overturned the city's total ban on residents being allowed to carry firearms outside their home **in a landmark decision for gun-rights activists.** (OP)
- (4) In Medical First, a Baby With HIV Is Deemed Cured (H)
Doctors announced on Sunday that a baby had been cured of an HIV infection for the first time, a startling development **that could change how infected newborns are treated and sharply reduce the number of children living with the virus that causes AIDS.** (OP)

Descriptions of social media impact are an interesting sub-category where both Impact and Superlativeness are constructed:

- (5) A video of the passionate haka performed by the comrades of three fallen New Zealand soldiers has gone viral, with **tens of thousands of people around the world** watching the clip.
- (6) Stephen Hawking's decision to boycott the Israeli president's conference has gone viral. **Over 100,000 Facebook shares** of the Guardian report at last count.

Impact here relates to uptake by news and social media. The answer to the always present putative audience query 'Why is this newsworthy?' is, in this case, that it has already captured massive global audience attention.

4.4 Negativity/positivity

Pointers to Negativity from the frequency lists include the noun *police* and the verbs *kill* and *die*. These are examples of lexis relating to crime and death. The verb *help* can also construct Negativity when it is used in connection with human suffering (where help is needed), but as suggested above it can also construct positive Impact. *Group* may establish Negativity together with co-occurring negative words (e.g., *extremist groups*).

The word form *good*, which occurs across twenty-one different news items, may appear to construct an event as positive on account of its evaluative meaning. However, concordances show that of the twenty-five instances, many do not construct the reported event as positive but have other meanings (e.g., *a good number of, good morning, good food, good-natured, be a good student, has any good idea about*), including some where positivity is negated, contrasted with alternative views or counter-factual:

- (7) [Item about prison inmates put on bread and water for destroying US flag]
It's just another vindictive policy that has nothing to do with running a **good** jail system.

- (8) [Item about planned ban on large sugary drinks]
I think it's a **good** idea,' said Sara Gochenauer, 21, a personal assistant from the Upper West Side. [...] But others said consumers should be free to choose.
- (9) [Item about review of animals and plants]
This update offers both **good** and **bad** news on the status of many species around the world.
- (10) [Item about survey that shows that public opinion often deviates from facts]
How can you develop **good** policy when public perceptions can be so out of kilter with the evidence?

The word *good* may also occur in research contexts in relation to benefits, where the research news itself is evaluated neutrally or positively/negatively, for example:

- (11) It sounds too **good** to be true but new research says having dessert – along with the traditional fry up – burns off the pounds. [research news presented as positive]
- (12) Bad news for dads: Babies 'should share mother's bed until age three' because it's **good** for their hearts [...] The controversial advice comes from [...]
[research news presented as controversial/negative]

In sum, there are only a few instances where the word *good* clearly seems to contribute to constructing Positivity, for instance in an announcement of an upcoming band tour, and in an item about teenagers saving a 5-year-old girl from a kidnapper:

- (13) Kiss and Def Leppard will team up this summer for a 42-city North American tour that will 'deliver **good** news and excitement,' says Kiss guitarist Paul Stanley.
- (14) Although the suspect remains at large, Tamar feels **good** about finding Jocelyn. 'I just feel like I did something very accomplishing today,' the teen said. The girl's family couldn't agree more.

In relation to semantic tags, Table 4 lists candidates that may be associated with Negativity, relating to events/issues such as disease, crime, death, warfare, violence, damage, and bad weather, as well as what USAS calls Evaluation: Good. The examples include negative lexis and evaluative language.¹⁵ As with the lemmas and word forms, the caveat is that it is unclear

¹⁵ The USAS category of Evaluation includes 'evaluative terms' depicting quality, truth, accuracy and authenticity (Archer *et al.*, 2002: 5–6), but is not further explained. The difference between negative evaluative language, reference to negative emotion/attitude and negative lexis in Table 1 is that evaluative language expresses opinion, while emotion

Negativity	B2- (Disease)	<i>injuries, wounds, disease, hurt, disabilities, diabetes, coma, stroke, cancer, leukemia...</i>
	G2.1- (Terms relating to crime/criminal activities)	<i>terrorism, criminal, fraud, kidnapping, illegal, guilty, crimes...</i>
	L1- (Terms relating to death)	<i>death, murder, suicide, killed, died, dead, executed, assassinated, slaughtered...</i>
	G3 (Warfare, defence and the army; weapons)	<i>shooting, gun, firearm, weapon, rocket, shot, rifle, military, invasion, war, missiles, ceasefire...</i>
	E3- (Violent/Angry)	<i>hit, wallop, angry, violently, assault, furious, attack, cruel, riot, violence...</i>
	A1.1.2 (Damaging and destroying)	<i>accident, destroyed, damages, violated, damaged, wreaked, harm, ruins, wreckage, break, crash, slash...</i>
	W4 (Weather)	<i>hurricane, snowstorm, flood, tornado, storm, breezy, monsoon, rains, wind, flooding...</i>
	G2.1 (Law and order)	<i>rules, police, law, court, regulations, legal, judge...</i>
Negativity or Positivity	S8+ (Helping)	<i>rescue, supporting, welfare, blessing, promoting, helps, benefit, aid...</i>
Positivity	A5.1+ (Evaluation : Good)	<i>great, good, super, wonderful, OK, well, positive, fine, fantastic...</i>

Table 4: Semantic tag candidates for Negativity/Positivity.

if these are in fact used to construct Negativity or Positivity, though the sheer wealth of negative tags seems to suggest that Negativity is a more important news value than Positivity in the corpus.

This assumption is confirmed by the manual analysis of headlines and OPs: only seventeen headlines and twenty-one OPs construct Positivity, while fifty-five headlines and fifty-one OPs establish Negativity, and twenty-seven headlines and OPs have unclear or no valence.

references label emotional experiences, and negative lexis concerns the use of vocabulary to describe negative events without automatically indicating writer dis/approval (see further Bednarek and Caple, 2012a, forthcoming).

4.5 Personalisation

From the frequency lists, the nouns *people*, *child*, *man*, *woman*, *family* and *parent* may be classified as pointers to Personalisation, if we assume that they are used to refer to ordinary people rather than elites. The frequency of lemmas pointing to children and families (CHILD, FAMILY and PARENT) is noteworthy here. The relevant semantic tags in the top 100 associated with Personalisation are S4f (Kin, female; e.g., *mother*, *mummy*, *bride*, *wives*, *mothers*, *sister* and *niece*), S2 tags (People; e.g., *people*, *children*, *females*, *man*, *child*, *woman*, *boy* and *girl*) and Z1 tags (Personal names; e.g., *Jim*, *George*, *Amy* and *Kathryn*). However, if these lemmas and word forms are used in a generic sense or to refer to groups, it is debatable whether Personalisation is constructed. One could argue that any reference to ordinary humans, as in the headline of Example 15, is less abstract and more personalised than text where they are not mentioned at all; but this is a very weak kind of Personalisation compared to the opening paragraph of Example 15, which focusses on one individual and his family.

- (15) Hundreds of kidnapped **Nigerian school girls** reportedly sold as brides to militants for \$12, **relatives** say (H)
 Samson Dawah was nervous. For two weeks, he had waited for any bit of information regarding his niece, who was among the 234 Nigerian school girls likely kidnapped by the terrorist group Boko Haram. This week, he gathered his extended family. He had news but also an unusual request. He asked that the elderly not attend. He wasn't sure they could bear what he had to say. (OP)

There are also instances where non-elite people are described as engaging in criminal activity (e.g., '**Drug addict parents** gave 23-month-old son methadone "like Calpol" before he died of overdose'), which makes them less 'ordinary' and more like 'criminals'. In other words, the news actors are constructed as 'other' rather than as 'us'. Both generic/group references and references to the criminal activities of non-elites could, therefore, be regarded as constructing only weak Personalisation (if they do so at all), but make up a considerable number of the references to non-elites in Hs and OPs (Table 5).

4.6 Proximity

Semantic tags and lemmas that may point to the construction of Proximity include the tags Z2 (Geographical names; e.g., *Washington*, *Salzburg*, *Afghanistan*, *American*, *Irish* and *Israeli*), M7 (Places; e.g., *place*, *country*, *town*, *area*, *region*, *district*, *province* and *hometown*) and the lemmas US/UNITED STATES, AMERICAN, STATE (N), CITY, AREA, COUNTRY. The latter construe Proximity if they are used to refer to a location close to the

News value	Sub-category	Headlines	OPs
Personalisation	Adults/mix/unspecified	12	14
	Children	10	5
	Total	22	19
Possible or 'weak' Personalisation	Generic or group	20	33
	Engaged in criminal activity	3	3
	Total	23	36
No Personalisation	N/A	54	44

Table 5: Personalisation.

audience. For example, US/UNITED STATES and AMERICAN would construct nearness for an American audience. With the nouns STATE, CITY, AREA and COUNTRY, the co-text and context would show whether or not they endophorically or exophorically refer to locations near the target audience. Similarly, it would be necessary to align geographical references tagged as Z2 and M7 with the news publication in which they occur to gain insights into the construction of Proximity for the original target audience of these publications. This is too time consuming to do for the whole corpus, as it includes publications from different countries targeted at different audiences (UK, USA, Australia and New Zealand).

However, the manual analysis of headlines/OPs suggests that fifteen headlines and thirty-one OPs construct Proximity for the original target audience by referring to the country or nationality of the news outlet, while thirty-four Hs/forty OPs do not. An additional fifty headlines and twenty-eight OPs 'possibly' construct Proximity, since they either refer to a country that is geographically or culturally close to the target audience or contain a cultural reference of some kind – such as *prom*, *Obama*, *gun-rights* and *haka*. Note that this analysis of the construction of Proximity for the original target audience ignores the audience of Facebook users who shared these items, for which Proximity may or may not be established (our data collection does not tell us who these users are and where they are located).

4.7 Superlativeness

Table 6 summarises semantic tag candidates and items from the frequency lists that point to Superlativeness—that is to say, various intensifiers, quantifiers (including numerals), focussing subjuncts, comparison, the noun

Lemmas, word forms, <i>n</i> -grams	Semantic tag and paraphrase	Some instances from the corpus
GROUP	S5 tags (Groups and affiliation)	<i>crowd, group, member, team, network, mob, federal, independent, affiliate...</i>
#, MORE, ONE, ALL, TWO, JUST, THAN, SO, ONLY, THREE, LIKE, WORLD, EVEN, FOUR, MANY, HIGH, VERY, SUCH, LEAST, EVERY, MUCH, LARGE, LONG, NUMBER, <i>five, six, really, several, second, various n</i> -grams with numbers (# symbol); <i>more than; at least; the world</i>	N1 (Numbers)	<i>two, thousands, 2004, million, 1974, 45...</i>
	N5 tags (Quantities, including N5+, N5++)	<i>percent, handful, dose, set, dozen, several, some, both, any, amount, number... many, dozens, onslaught, hundreds, multiple, enough, much... increased, more, most, a lot, as well as, extra, added...</i>
	N5.1+ (Terms depicting maximal/maximum quantities)	<i>all, any, every, each, full, total...</i>
	A13.3 (Degree: Boosters Intensifiers that amplify to a high degree)	<i>very, highly, helluva-really, particularly, increasingly, greatly, extremely, considerably...</i>
	A14 (Focussing subjuncts that draw attention to/focus upon X.)	<i>just, only, solely, alone, especially, exclusively...</i>
	A13.2 (Intensifiers that amplify to the upper extreme)	<i>most, fully, absolutely, literally, perfectly, absolutely, totally...</i>

Table 6: Semantic tag candidates for Superlativeness.

world and the bigram *the world*.¹⁶ The lemma *group* and its co-text can establish Superlativeness, when used to emphasise quantity (e.g., ‘a huge group of volunteers’). *Least, at least* and *more than* are also included on the assumption that they can construct an ensuing number as high. Concordancing of the lemma LEAST shows that of forty-two occurrences, thirty-nine are instances of *at least* and most of these are followed by a

¹⁶ In including all these items here I have been less conservative than with other news values, as some of these can function in different ways (for instance, *so* can be a conjunction as well as an intensifier).

t's Ural mountains has injured at least 950 people, as the shoes were dead. It looked to us at least 70 per cent of the image top, inspect minutely and ask at least 6 questions about every 1000 km's atmosphere at a speed of at least 54,000 km/h (33,000 mph). Repeat everything you say at least 5 times. Test 8: Grocery shopping nothing else on television for at least 5 years. Test 11: Messengers suffered cuts and bruises but at least 46 remain in hospital. A predicted path. The quake killed at least 222 people, injured near employees, provided they work at least 20 hours a week and have a disability is anathema, has killed at least 2,300 people since 2010.

Figure 1: Some concordances for *at least*

number, with many (but not all, compare Figure 1) constructing the reported event as of high intensity or large scope/scale. Sometimes Superlativeness is combined with Impact and Negativity in instances that establish the negative consequences for a large number of people (e.g., 'has injured at least 950 people' and 'killed at least 222 people').

As with most other pointers, further qualitative investigation would be necessary. While it is not possible to do this for the whole corpus, we can have a brief look here at some pointers that occur across at least four OPs: Examination of all thirty-five occurrences of numbers (#) in the OPs shows that some of them do construct Superlativeness (especially those expressing large amount), while others refer to date and time, the age of news actors, or other aspects (e.g., '22-caliber rifle'). The word forms *world*, *thousands* and *most* construe Superlativeness in several OPs (but with exceptions; e.g., 'planned to travel the world'), while the word *one* does so only on occasion. Different phraseologies can be identified:

- vague large number (e.g., *thousands of*) + [ordinary] NEWS ACTOR);
- [ordinary] NEWS ACTOR *around the world*;
- *most*, *-est* [+SCOPE] ('our embattled PM's **most** embarrassing moments', 'in the **most** ambitious effort yet', 'the **oldest in the world**' and 'one of the **strongest** storms recorded **on the planet**').

These and other phraseologies are used in the corpus to construct the uniqueness of a news actor ('the oldest in the world'), emphasise the large amount of non-elite news actors involved ('thousands of NEWS ACTOR' and 'NEWS ACTOR around the world'), the degree of an event or action ('**one**

of the strongest storms recorded on the planet; our embattled PM's **most** embarrassing moments'), or the scope of a discovery ('a whole new **world** of ... has come to light'). However, the manual analysis of headlines and OPs shows that not all construct Superlativeness: fifty-nine headlines and forty-two OPs do not establish Superlativeness, while forty headlines and fifty-four OPs do (with three additional 'possible' cases in OPs).

4.8 Timeliness

Several items from the frequency lists could be used to establish Timeliness (including newness): this includes lemmas that point to constructions of events as recent (LAST and WEEK),¹⁷ immediate (NOW and LATE [includes *latest*]), about to happen (*soon*), continuing (STILL) or starting (the verb START), or as the first (FIRST and 'the first') or otherwise new (NEW, NEWS). In fact, when considering the lemmas NEW and NEWS, the former is a much better predictor of the construction of Timeliness than the latter (see Bednarek and Caple, forthcoming) – unless it is part of a proper noun (*New York* and *New Zealand*). In contrast, the lemma NEWS occurs most often in references to news organisations (e.g., 'told Fox News' and 'according to Dutch News'). It can refer to new information and construct news value ('latest news') and it is interesting to note that it is in some instances accompanied by evaluative language which may establish additional news value: 'good and bad news', 'heart-breaking news', 'dangerous news', 'good news and excitement', 'the great news is' and 'the news got worse'. However, on the whole, NEWS is not a good predictor of the construction of newsworthiness.

Semantic tags that seem like potential candidates include T1.3 (Time: Period; e.g., *days, hours, years, months, night, week, period, evening, Monday* and *May*), T 1.1.2 (Time: General: Present; simultaneous; e.g., *now, today, yet, instant, present, updates, meanwhile, instantaneous, currently, tonight, daily* and *coincides with*), T1.1.3 (Time: General: Future; e.g., *will, future, tomorrow, impending, coming, gonna, shall* and *soon*), T2 tags (Time: Beginning and ending; e.g., *start, remain, continue, former, beginning, source, remain, established, initial, stopped, still* and *going on*), T3 tags (Time: Old, new and young; age; e.g., *adult/s, x-old, young, baby, recent, latest, original, of this age, middle-aged, new, over the age of* and *death*) and A10+ (Finding, showing; e.g., *revelation, finding, showing, exposed, revealed, traced, indicated, found* and *debunked*). Again, we can see that semantic tags are at times too broad to be useful for DNVA. For example, 'Time: Old, new and young; age' includes items relevant to the construction

¹⁷ I categorised WEEK as a pointer to Timeliness but not YEAR or MONTH, on the assumption that references to *last/this year/month*, for example, would in most cases refer to a point in time too far removed from the time of publication to emphasise the Timeliness of the constructed event.

of Timeliness such as *recent*, *latest* and *new*, and irrelevant items such as *years old* and *age group*.

The manual analysis of Hs and OPs suggests that about 70 to 80 percent of Hs/OPs construct Timeliness (the numbers are even higher, especially in the Hs, if ‘possible’ instances are included and ‘not applicable’ instances are excluded). In view of the corpus design itself, I expected the news value of Timeliness to be constructed in the data, since our working definition of ‘news’ items included that the reported event had to be new or a new development. In other words, we cannot independently explore whether or not news stories constructed as ‘timely’ are shared more than news stories that are not constructed as timely. It is also worth mentioning that news items might in theory be shared on social media long after their publication date, so that Timeliness might not be established for the audience of shared news, even if it was established as such for the original target audience.

4.9 Unexpectedness

I have not identified any semantic tags in the top 100 most frequent USAS tags that point to Unexpectedness, although the lemma *FIRST* and the bigram *the first* do—when they construct an event as unusual (e.g., ‘the first time since’) rather than new (Timeliness). In fact, the manual analysis of Hs and OPs shows that although some explicit evaluations occur (e.g., *unusual*, *bizarre*, *rare*, *miracle* and *startling*) Unexpectedness is often constructed by relatively factual descriptions of events that many would evaluate as unexpected, as in Example 16.

- (16) The Penguin foundation has a global callout for knitters to make pullovers for penguins in rehab.

It is a reasonable assumption that most readers would be surprised that (a) penguins can be in ‘rehab’ and (b) that they would need or wear pullovers. Often it seems to be research findings that are constructed as both new and unexpected (e.g., ‘new research says having dessert—along with the traditional fry up—burns off the pounds’). In total, the construction of Unexpectedness is a feature of forty-one headlines and twenty-six OPs and ‘possibly’ of a further forty-eight OPs and forty-one Hs.

4.10 Quantitative summary of results and methodological reflection

Tables 7 and 8 (pp. 250 and 251, respectively) provide a quantitative summary of the results. Regarding the coding of the Hs and OPs for all ninety-nine items, it must be stressed that these numbers represent trends rather than facts, since researcher subjectivity plays a role. With respect to the potential ‘pointers’ in the frequency lists and tags, their identification

News values	'Yes'		'No'		'Possible'		Potential pointers (word forms, lemmas and <i>n</i> -grams with a range of at least twenty)	Potential pointers (among top 100 most frequent tags)
	H	OP	H	OP	H	OP		
Consonance	–	–	79	72	20	27	no potential pointers	<i>Full text corpus</i> A6.2+ (Comparing: Usual)
Eliteness	44	49	35	34	20	16	OFFICIAL, UNIVERSITY, GOVERNMENT, PRESIDENT, US/UNITED STATES, STATE (noun), AMERICAN (CITY, AREA, COUNTRY, GROUP)	G1.1 tags (Government, <i>etc.</i>); S7.1+ (Terms depicting power/authority/influence and organisation/administration); G2.1 (Law and order); Z2 (Geographical names); M7 (Places); S5 tags (Groups and affiliation); Z3c (Other proper names); Z1 tags (Personal names)
Impact	12	18	86	78	1	3	RESULT, LEAVE, CAUSE, BECOME, FOLLOW, HELP	A11.1+ (Abstract terms denoting importance/significance); A2.2 (Affect: Cause/Connected); A2.1+ (General/abstract terms denoting [propensity for] change); S8+ (Helping)
Personalisation	22	19	54	44	23	36	PEOPLE, CHILD, MAN, WOMAN, FAMILY, PARENT	S4f (Kin, female), S2 tags (People); Z1 tags (Personal names)
Proximity	15	31	34	40	50	28	US/UNITED STATES, AMERICAN (STATE, CITY, AREA, COUNTRY)	Z2 (Geographical names); M7 (Places)
Superlativeness	40	54	59	42	–	3	#, MORE, ONE, ALL, TWO, JUST, THAN, SO, ONLY, THREE, LIKE, WORLD, EVEN, FOUR, MANY, HIGH, VERY, SUCH, LEAST, EVERY, MUCH, LARGE, LONG, NUMBER, <i>five, six, really, several, second</i> , various <i>n</i> -grams with numbers (# symbol); <i>more than; at least; the world</i> (GROUP)	N1 (Numbers); N5 tags (Quantities); N5.1+ (maximal/maximum quantities); A13.3 (Degree: Boosters); A14 (Focussing subjuncts); A13.2 (Intensifiers); S5 tags (Groups and affiliation)
Timeliness	69	78	8 (+16 N/A)	17 (+4 N/A)	6	–	LAST, WEEK, NOW, LATE [incl. <i>lates[]</i>], <i>soon</i> , STILL, START (V), FIRST, <i>the first</i> , NEW (NEWS)	T1.3 (Time: Period); T 1.1.2 (Time: General: Present; simultaneous); T1.1.3 (Time: General: Future); T2 tags (Time: Beginning and ending); T3 tags (Time: Old, new and young); A10+ (Finding, showing)
Unexpectedness	41	26	17	25	41	48	FIRST, <i>the first</i>	–

Table 7: Quantitative trends and potential pointers to news values.

	H	OP	Full text corpus: potential pointers (word forms, lemmas and <i>n</i> -grams with a range of at least twenty)	Full text corpus: potential pointers (among top 100 most frequent tags)
Negativity	55	51	POLICE; KILL; DIE (GROUP)	B2- (Disease); G2.1- (Terms relating to crime/criminal activities); L1- (Terms relating to death); G3 (Warfare, defence and the army; weapons); E3- (Violent/Angry); A1.1.2 (Damaging and destroying); W4 (Weather); G2.1 (Law and order)
Positivity	17	21	(<i>good</i>)	A5.1+ (Evaluation: Good)
Unclear or no valence	27	27	(HELP)	S8+ (Helping)

Table 8: Quantitative trends and potential pointers to Negativity and Positivity.

was based on meaning potential rather than actual use since comprehensive concordancing could only be undertaken in some cases (compare Sections 3 and 4).¹⁸ As these tables illustrate, there is no unequivocal correlation between the frequency of Hs/OPs that construct a specific news value and the frequency of potential pointers. For instance, there are few potential pointers to Unexpectedness, yet this news value is constructed in at least forty-one (of ninety-nine) corpus items. My hypothesis is that some news values, such as Unexpectedness and Consonance, are perhaps more frequently established through implicit means in shared news, including relatively factual descriptions of events—as we have seen with Unexpectedness. The frequency of pointers also depends on differences between linguistic resources. For example, the fact that more items are included under Superlativeness does not in itself indicate that the news value is more important than others—grammatical items such as intensifiers and quantifiers simply tend to be higher in frequency and range.

In general, all ‘traditional’ news values appear to be construed in shared news but there is a fair amount of variety. For instance, it is not the case that Facebook users *only* share items about news actors constructed as ‘elite’ (e.g., Hollywood stars) or about news constructed as ‘positive’ (e.g., ‘feel-good’ stories). This variety might reflect the fact that Facebook users, like other online news consumers, consist of different types with varying interests and behaviours (see, for example, Olmstead *et al.*, 2011). On the other hand, the news values of Eliteness, Superlativeness, Unexpectedness, Negativity and Timeliness seem especially important in shared news, although the latter may result from the corpus design as explained above.¹⁹

As far as the corpus techniques of word lists and semantic tagging are concerned, some key limitations have emerged. First, much further qualitative analysis is necessary for a comprehensive picture, because not all lemmas/word forms are good predictors for newsworthiness construction (see my analysis of *good* and *news*) and a semantic tagger cannot be used as a news values tagger since the semantic fields are often too broad. Second, news values may not necessarily be established by frequently recurring word forms or words from the same semantic field (see Consonance). Unless a specific news values tagger is developed, qualitative analysis is, thus, crucial—this can be undertaken through concordancing or through manual annotation based on close reading.

5. Conclusion

To conclude, we can revisit some of the questions that have arisen about shared news: are surprising/unexpected, positive and affective/emotional

¹⁸ Brackets have been used in the ‘word form/lemma/*n*-grams’ column to identify questionable pointers, which certainly only construe the respective news value in certain co-texts.

¹⁹ Proximity also emerges as important if cultural references are included (see Bednarek and Caple, forthcoming).

news items shared more? This study suggests that unexpected news is in fact an important sub-category of socially shared news from heritage news sources, not just, but also in relation to, research news. Research findings that are shared widely tend to be constructed both as new and unexpected. This is a useful insight for different user groups who desire a wider impact (e.g., in the health sector, and NGOs and academics). In relation to positivity, it seems that Facebook users share a mix of positive and negative items from heritage news media, although Negativity seems to be a more important news value than Positivity, in line with Hansen *et al.*'s (2011) study of re-tweeted news. Some of these items concern controversial topics (e.g., gun rights and Israel) and some construct the event as high in negative impact (e.g., killing or injuring many people). Together with the fact that children are often mentioned and that surprising news is important, it is likely that affective news is shared more, since these sub-categories are likely to evoke audience emotions.

What are the lessons for corpus linguistic research? In relation to DNVA we are faced with many of the same problems as corpus linguists who research evaluation more generally (Hunston, 2011; and Potts *et al.*, 2015: 168–9), including the scalar nature of news values. Two of the challenges consist of the multiple meanings and uses of words, and the broad nature of semantic fields. This makes it necessary to combine quantitative with qualitative analysis, but even with a small corpus, a truly comprehensive analysis is too time-consuming. A partial picture may be provided through selective qualitative corpus analysis (as in Potts *et al.*, 2015) or through discourse analysis of one important structural element of each corpus item—in this case, the headline and opening paragraph. However, it is important to remain aware of genre differences, which result in different types of Hs and OPs.

Another venue for a corpus project concerns the development of a news values tagger. For example, it might be possible to identify a large group of words and expressions that are both frequent in news discourse and at the same time good predictors of news values, such as *government, president, died, killed, new, at least*, vague large numbers (e.g., *thousands*), and *around/in the world*. It might also be possible to compile a list of noteworthy politicians, stars, sports people, *etc.*, and match proper names against that list to identify Eliteness. When working with corpora containing news outlets from only one country, it should also be feasible to identify references to geographical locations in that country to identify Proximity.

In general, corpus linguistic analysis of evaluation can benefit from taking into account the existence of context-sensitive goals and values. Both goals and values are arguably tied very much to the cultural context of the products (texts) that corpus linguists study, including the people who produce these products, the process by which these products are produced, and the professional context in which this occurs. This presents a challenge for corpus linguists who analyse evaluation in finished products, but tend to

pay less attention to the other three ‘Ps’. One way of considering these is by taking into account the value systems of particular professions, for example ‘news values’.

Acknowledgments

This paper is an output of the Australian Research Council Linkage Project grant ‘Sharing News Online: Analysing the Significance of a Social Media Phenomenon’ (LP 140100148). I am grateful to the project members Tim Dwyer, Fiona Martin and James Curran as well as Andrew Hunter, Hal Crawford and Domagoj Filipovic from our industry partners Share Wars and Mi9 for helpful discussions and access to ShareWars’ Likeable Engine. Tim Dwyer and Fiona Martin also provided helpful comments on this paper. Finally, I wish to thank two research assistants, Joel Nothman, who compiled all the verbal text and acted as expert Data Mining consultant, and Penelope Thomas, who corrected this data manually.

References

- Anderson, M. and A. Caumont. 2014. ‘How social media is reshaping news’, Pew Research Center Fact Tank, 24 September 2014. Accessed 10 April 2015 at: <http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/>
- Archer, D., A. Wilson and P. Rayson. 2002. ‘Introduction to the USAS Category System’. Available online at: <http://ucrel.lancs.ac.uk/usas/usasguide.pdf>
- Bednarek, M. 2006. *Evaluation in Media Discourse. Analysis of a Newspaper Corpus*. London and New York: Continuum.
- Bednarek, M. 2015. ‘Coding manual’. Available online at: <http://www.newsvaluesanalysis.com>
- Bednarek, M. 2016. ‘Voices and values in the news: news media talk, news values and attribution’, *Discourse, Context, Media* 11, pp. 27–37.
- Bednarek, M. and H. Caple. 2012a. *News Discourse*. London and New York: Continuum.
- Bednarek, M. and H. Caple. 2012b. “‘Value added’”: Language, image and news values’, *Discourse, Context, Media* 1, pp 103–13.
- Bednarek, M. and H. Caple. 2014. ‘Why do news values matter? Towards a new methodological framework for analyzing news discourse in Critical Discourse Analysis and beyond’, *Discourse and Society* 25 (2), pp 135–58.

- Bednarek, M. and H. Caple. Forthcoming. *The Discourse of News Values: How News Organisations Create Newsworthiness*. Oxford and New York: Oxford University Press.
- Bell, A. 1991. *The Language of News Media*. Oxford: Blackwell.
- Bell, E. 2015. 'The 12th annual Hugh Cudlipp lecture', *The Guardian*, 28 January 2015. Available online at: http://www.theguardian.com/media/2015/jan/28/emily-bells-2015-hugh-cudlipp-lecture-full-text?CMP=share_btn_link
- Berger, J. 2011. 'Arousal increases social transmission of information', *Psychological Science* 22 (7), pp. 891–3.
- Berger, J. and K.L. Milkman. 2011. 'What makes online content viral', *Journal of Marketing Research*, Ahead of Print, DOI: 10.1509/jmr.10.0353, pp 1–17.
- Bruns, A., T. Highfield and S. Harrington. 2013. 'Sharing the news: Dissemination of links to Australian news sites on Twitter' in J. Gordon, P. Rowinski and G. Stewart (eds) *Br(e)aking the news. Journalism, Politics and New Media*, pp. 181–209. Bern and New York: Peter Lang.
- Caple, H. 2013. *Photojournalism: A Social Semiotic Approach*. Basingstoke and New York: Palgrave Macmillan.
- Caple, H. and M. Bednarek. 2013. 'Delving into the Discourse: Approaches to News Values in Journalism Studies and Beyond', Oxford: Reuters Institute for the Study of Journalism, The University of Oxford. Available online at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Delving%20into%20the%20Discourse_0.pdf
- Cotter, C. 2010. *News Talk: Investigating the Language of Journalism*. Cambridge: Cambridge University Press.
- Crawford, H., D. Filipovic and A. Hunter. 2015. *All Your Friends Like This: How Social Networks Took Over News*. Harper Collins.
- van Dijk, T. 1988. *News as Discourse*. Hillsdale, NJ: Erlbaum.
- Feez, S., R. Iedema and P.R.R. White. 2008. *Media Literacy*. Surry Hills, NSW: NSW Adult Migrant Education Service.
- Galtung, J. and M.H. Ruge. 1965. 'The structure of foreign news', *Journal of Peace Research* 2 (1), pp. 64–91.
- Hansen, L.K., A. Arvidsson, F.Å. Nielsen, E. Colleoni and M. Etter. 2011. 'Good Friends, Bad News: Affect and Virality in Twitter'. Paper presented at The 2011 International Workshop on Social Computing, Network, and Services (SocialComNet 2011), Loutraki, Crete, Greece.
- Hermida, A., F. Fletcher, D. Korell and D. Logan. 2012. 'Share, like, recommend', *Journalism Studies* 13 (5), pp. 815–24.

- Hood, S. 2010. *Appraising Research: Evaluation in Academic Writing*, London: Palgrave Macmillan.
- Hunston, S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. New York and Oxon: Routledge.
- Iedema, R., S. Feez and P. White. 1994. *Media Literacy*. Sydney: Disadvantaged Schools Program, N.S.W. Department of School Education.
- Lee, C.S. and L. Ma. 2012. 'News sharing in social media: the effect of gratifications and prior experience'. *Computers in Human Behavior* 28 (2), pp 331–9.
- Mahlberg, M. and M.B. O'Donnell. 2008. 'A fresh view of the structure of hard news stories' in S. Neumann and E. Steiner (eds) *Online Proceedings of the 19th European Systemic Functional Linguistics Conference and Workshop*, Saarbrücken, 23–25 July 2007. Available online at: <http://scidok.sulb.uni-saarland.de/volltexte/2008/1700>
- Martin, F. and T. Dwyer. 2015. 'How Did They Get Here? The Likeable Engine, Dark Referrals and the Problematic of Social Media News Analytics'. Paper presented at *IAMCR 2015*, Montreal, 12–16 July 2015.
- Newman, Nic 2011. 'Mainstream media and the distribution of news in the age of social discovery', University of Oxford: Reuters Institute for the Study of Journalism, September 2011. Accessed 10 April 2015 at: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Mainstream%20media%20and%20the%20distribution%20of%20news%20in%20the%20age%20of%20social%20discovery.pdf>
- O'Donnell, M. 2015. UAM Corpus Tool. Version 3.2. Available online at: <http://www.wagsoft.com/CorpusTool/download.html>
- Olmstead, K., A. Mitchell and T. Rosenstiel. 2011. 'Navigating news online: where people go, how they get there and what lures them away', Pew's Research Center's Project for Excellence in Journalism. Accessed 25 August 2013 at: http://www.journalism.org/analysis_report/navigating_news_online
- Pflaeging, J. 2015. "'Things that matter, pass them on": ListSite as viral online genre', *10plus1: Advancements in Linguistics* 1, pp. 156–81. Accessed 7 December 2015 at: http://10plus1journal.com/wp-content/uploads/2015/09/12_JOU_ART_Pflaeging.pdf
- Potts, A. 2013. *At Arm's Length: Methods of Investigating Constructions of the 'Other' in American Disaster and Disease Reporting*. PhD Thesis. Lancaster University, UK.
- Potts, A., M. Bednarek and H. Caple. 2015. 'How can computer-based methods help researchers to investigate news values in large datasets? A corpus linguistic study of the construction of newsworthiness in the

- reporting on Hurricane Katrina', *Discourse and Communication* 9 (2), pp. 149–72.
- Reuters Institute Digital News Report. 2014. Oxford: Reuters Institute for the Study of Journalism, University of Oxford. Accessed 10 April 2015 at: <http://bit.ly/digitalnews2014>
- Sinclair, J.M. 2004. 'Corpus and text: basic principles' in M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books/Arts and Humanities Data Service, pp. 1–16.
- Thompson, G. and S. Hunston. 2000. 'Evaluation: An introduction' in S. Hunston and G. Thompson (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse*, pp 1–27. Oxford: Oxford University Press.
- Wasike, B.S. 2013. 'Framing news in 140 characters: How social media editors frame the news and interact with audiences via Twitter', *Global Media Journal – Canadian Edition* 6 (1), pp. 5–23.
- Weeks, B.E. and R.L. Holbert. 2013. 'Predicting dissemination of news content in social media: A focus on reception, friending and partisanship', *Journalism and Mass Communication Quarterly* 90 (2), pp. 212–32.
- White, P.R.R. 1997. 'Death, disruption and the moral order: the narrative impulse in mass media "hard news" reporting' in F. Christie and J.R. Martin (eds) *Genres and Institutions: Social Processes in the Workplace and School*, pp. 101–33. London: Cassell.