

Coding manual for linguistic analysis (using UAM Corpus Tool) in:

- *The Discourse of News Values*, OUP, 2017 [Bednarek & Caple]
- ‘Investigating evaluation and news values in news items that are shared via social media’, *Corpora* [Bednarek, 2016]

Cite as:

Bednarek, M. (2015). ‘Coding manual for linguistic analysis’, available at www.newsvaluesanalysis.com.

Important note:

This coding manual is supplementary material for the two publications listed above. The coding for the analysis described in these publications was based on the framework for linguistic analysis of news values presented in Chapter 4 of Bednarek & Caple (2017), taking into account co-text and context. This manual aims not to repeat too much information from that chapter. Generally, the coding therefore simply applied the framework from the book, but further information on **how** this framework was applied (including modifications/alterations) is provided in this manual (e.g. coding choices for special cases). It is **not** advisable to read or apply this coding manual without first having read Bednarek & Caple (2017).

The coding manual first re-prints the **inventory** of linguistic resources used to generally guide the analysis (pages 2-5), before providing further information on the **general procedure** (page 6) and more **specific coding decisions** (page 7 onwards).

The inventory can also be downloaded as a separate document from the DNVA website (www.newsvaluesanalysis.com). This inventory can be used to guide any DNVA research, while the remainder of the coding manual is supplied here primarily because of research ethics (transparency), rather than as a manual to be used in future research. It simply explains how the coding/analysis was undertaken for the project described in these two publications, rather than prescribing how any coding/analysis of news values should proceed.

References that are mentioned in the manual are listed at the end of this document (page 15).

Inventory of linguistic devices that often construct newsworthiness in English-language news (general guide to the analysis, as further discussed in Chapter 4 of Bednarek & Caple 2017)

News value	Linguistic Device	Examples
<i>Consonance</i> ([stereo]typical)	References to stereotypical attributes or preconceptions	<i>Drug addict parents gave 23-month-old son methadone 'like Calpol'; Most broody mothers see having a child as a wonderful gift from God; ... another eventful Australia Day, which most Australians enjoyed in the best way they knew how – laying under the sun on a white-sand beach or enjoying a barbecue with friends and family.</i>
	Assessments of expectedness/typicality	<i>a man whose love of luxury and lavish parties is legendary; ...as dire Diaz campaign ends in typical style</i>
	Similarity with past	<i>...as the US came to terms with yet another mass shooting; America is once again torn apart by race and police power</i>
	Explicit references to general knowledge/traditions	<i>In keeping with the Germans' well-known love of beer</i>
<i>Eliteness</i> (of high status or fame)	Various high status markers, including	
	role labels	<i>U.S. District Court Judge Scott Skavdahl; Professor Roger Stone; Snohomish county fire district 21 chief; celebrity chef Jamie Oliver; Abba legend Björn Ulvaeus</i>
	status-indicating adjectives	<i>the prestigious Man Booker prize; the city's top cop; a senior World Bank executive; a key federal government minister; long-term industry observers; well-placed government sources;</i>
	recognised names	<i>Hillary Clinton; Abba; the Olympics; the Oscars; Harvard university; the World Health Organisation</i>
	descriptions of achievement/fame	<i>Ronnie Barker of the Two Ronnies fame; two people who were selling millions of records a year; The Norwegian entertainer was also very popular in neighbouring countries</i>
	use by news actors/sources of specialised/technical terminology, high-status accent or sociolect [esp. in broadcast news]	N/A
<i>Impact</i> (having significant effects or consequences)	Assessments of significance	<i>a potentially momentous day; in a historic legal case; a crucial annual conference</i>
	Representation of actual or non-actual significant/relevant consequences, including abstract, material or mental effects	<i>note that will stun the world; leaving scenes of destruction; Millions of Australian homes and businesses could be hit with bigger phone and internet bills; thousands of people may be massacred</i>

Negativity/Positivity (negative/positive)	References to negative/positive emotion and attitude	<p><i>concerns</i> about even remote chances of Ebola exposure; <i>fury</i> as primary head takes week off in term to fly to Caribbean; a move that has outraged local politicians; amid signs of panic; ‘First hydrogen bomb test’ condemned;</p> <p>Pale but smiling, former U.S. Marine Amir Hekmati recounted Tuesday how disbelief turned to joy; Baltimore residents celebrate charges in Gray case</p>
	Negative/positive evaluative language	<p>Corbyn’s shambolic reshuffle; a violent thug who had no interest in Islam; shoddy financial advice;</p> <p>the brilliant astrophysicist; the perfect end to another eventful Australia Day</p>
	Negative/positive lexis	<p>Boy, 8, one of 3 killed in bombings at Boston Marathon; scores wounded; Western black rhino declared extinct; Flint residents protest high bills for ‘poison’ water; 13 migrants drown as boat capsizes off Malaysia;</p> <p>Nigeria has been declared officially free of Ebola; Teens chase kidnapping suspect on bikes, save 5-year-old girl; ... a baby with HIV is deemed cured</p>
	Descriptions of negative (e.g. norm-breaking) or positive behaviour	<p>Hospitals don’t have enough beds, and there aren’t enough ambulances; Treasurer Joe Hockey has broken his promise to balance the budget by 2019;</p> <p>[Canadian Prime Minister] Trudeau, who last year unveiled a cabinet with an equal number of men and women “because it’s 2015”</p>
Personalisation (having a personal/human face)	References to ‘ordinary’ people, their emotions, experiences	<p>Mike’s devastated owner; 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’</p>
	Use by news actors/sources of ‘everyday’ spoken language, accent, sociolect [esp. in broadcast news]	N/A
Proximity (geographically or culturally near) [continued on next page]	Explicit references to place or nationality near the target community	<p>A federal judge in the District of Columbia ...; A skeleton found beneath a Leicester car park...; A Texas father caught a man sexually assaulting his 4-year-old daughter; Australian nurse in Ebola scare;</p>

<i>Proximity</i> (geographically or culturally near) <i>[continued from previous page]</i>	References to the nation/community via deictics, generic place references, adjectives	<i>Homegrown terrorist Mohamed Elomar pledges to bring the horror here; a potential attack on the nation's capital and the country's highest office</i>
	Inclusive first person plural pronouns	<i>Red alert over the plot to attack our nation's leaders Is this the end of our local newsagents?</i>
	Use by news actors/sources of [geographical] accent/dialect [esp. in broadcast news]	N/A
	Cultural references	<i>Teenager takes his great-grandmother to prom; Soldiers' farewell haka footage goes viral</i>
<i>Superlativeness</i> (of high intensity/large scope)	Intensifiers	<i>The Ebola outbreak [...] will get significantly worse; A sensational corruption inquiry has concluded...; The US government says it is "deeply concerned" by reports that...</i>
	Quantifiers	<i>hundreds who flew with an infected nurse; the country's two-week-old political crisis; a tragedy of epic proportions; a ... \$356 million loss</i>
	Intensified lexis	<i>U.S. forces hammered ISIS fighting positions, vehicles and buildings; Robbers smash display cases; Police seek motive in Idaho shooting rampage that killed 3; they were petrified</i>
	Metaphor and simile	<i>...country towns in northern NSW are battling a tsunami of crime; a June wildfire that ... ripped through as if the land had been doused with gasoline</i>
	Comparison	<i>Foxtons' stock price was rising faster than the cost of a Mayfair penthouse; Brad Pitt and Angelina Jolie's wedding was so secret Jolie's father Jon Voight did not know it had taken place; ...around 5,000 more suicides in Europe and North America; ... the largest drug ring in Detroit history; ... one of the world's most prolific serial killers; ... 2014 surpassed 2010 as the warmest year;</i>
	Repetition	<i>...with building after building flattened or punctured by shells</i>
	Lexis of growth	<i>The volume of email cloaked in encryption technology is rising; ...adding to a growing list of healthcare workers in West Africa hit by the epidemic; It had sheer scale, scope, the length and the breadth of the evil unfolded;</i>
	<i>Only/just/alone/already</i> + time/distance or related lexis	<i>Already this year 64 clandestine ice labs have been busted and dismantled; almost a hundred foreigners ... were arrested in one raid alone</i>

<i>Timeliness</i> (recent, ongoing, about to happen, new, current, seasonal)	Temporal references	<i>A terrorist attack...is now regarded as “likely”;</i> <i>Labour will today offer;</i> <i>yesterday’s flash flooding</i>
	Present and present perfect	<i>it is testing our emergency resources;</i> <i>INDONESIA’S Justice Ministry is about to isolate...;</i> <i>rescuers have been trying to pluck survivors</i>
	Implicit time references through lexis (e.g. <i>ongoing, under way, begin</i>)	<i>A murder investigation is under way in Dublin;</i> <i>Search ongoing for missing Victoria woman Karen Chetcuti</i>
	References to:	
	current trends	<i>“selfie” — the smartphone self-portrait — has been declared word of the year for 2013</i>
	seasonality	<i>... as Public Health England urged people to keep their homes well heated this winter;</i>
<i>Unexpectedness</i> (unexpected)	change/newness	<i>In an unexpected development;</i> <i>Bowser says change from GLBT to LGTB is ‘in keeping with the mainstream vocabulary.’;</i> <i>Mint 1969 Shelby GT500 found under 40 years of dust;</i> <i>EU leaders pick new top diplomats</i>
	Evaluations of unexpectedness	<i>one of the strangest scandals;</i> <i>an unusual case in a city where prosecutions of police for excessive [sic] are rare</i>
	References to surprise/expectations	<i>shock at North Cottesloe quiz night;</i> <i>people just really can’t believe it</i>
	Comparisons that indicate unusuality	<i>Sydney’s wettest August in 16 years;</i> <i>the first time since 1958;</i> <i>I’ve lived in Toowoomba for 20 years and I’ve never seen anything like that</i>
	References to unusual happenings	<i>British man survives 15-storey plummet;</i> <i>German MPs considering a return to typewriters to combat spy activity;</i> <i>Queensland woman fights off kangaroo with backpack</i>

General procedure

1. The text was manually coded using UAM Corpus Tool (O'Donnell 2015).
2. Headline (H) and opening paragraph (OP) were coded separately: first all headlines were coded, then all opening paragraphs. Each H and each OP constituted a separate file or item to be coded.
3. Each file (H/OP) was coded for each news value in turn – for instance, all headlines were first coded for Timeliness, then for Consonance, and so on. This allowed a focus on one news value, resulting in more accurate, systematic and consistent analysis.
4. The first step was to code ten items (e.g. headlines) completely, which was used to set up the coding scheme. The remainder of the items was then coded, which resulted in some refinements to the original coding scheme (for example, sub-categories). Every time the coding scheme was changed, *all* files were re-coded from the beginning. Any reference to news organisations and journalists was left uncoded, as they are part of the newsmaking process rather than news actors.
5. The corpus was coded in two different ways: Firstly, each linguistic device or 'markable' (Taboada et al 2014: 133) – word, phrase, clause, sentence – was annotated taking into account its co-text and context (and using the inventory on pp. 2-5 as guide). With this coding procedure, one item (H/OP) may be coded several times for the same news value. For example, one headline may contain more than one device for Negativity and each device would then be annotated separately, meaning that one headline would be annotated several times for Negativity. This was useful for identifying common linguistic devices in the corpus and refining Chapter 4, but made quantification much more complex. Secondly, each headline and each OP was coded *once* per news value; here, the overall meaning of the whole unit was responsible for the selected coding. This coding was used for quantifying which news values were constructed in Hs and OPs, with only three options for most news values: 'yes' (clearly constructs respective news value), 'no' (clearly does not construct respective news value), 'possible' (may construct the respective news value to a small degree or a debatable instance or a special case in some other way). For all news values, it is the case that if a H/OP contains at least one device that constructs the respective news value it is coded as 'yes', even if it comprises other devices that only 'potentially' construct the same news value.
6. Although the files were primarily coded by one author, several steps were taken to improve the quality of the coding: A coding manual was set up for each news value to manage the consistency of coding; all problematic instances were marked and discussed by both authors to resolve their final coding. With Consonance and Unexpectedness, **all** instances (problematic or not) were discussed amongst the authors.¹ Once the headlines and OPs had been coded once, the coded items were left un-read for several weeks, and then all codings were checked again. At this stage, the codings of headlines and OPs were also compared against each other and checked for consistency.

¹ A note on inter-coder agreement: While it would have been possible to have several coders use the manual to code the items, and inter-coder agreement could then have been measured, this would only have provided information on the quality of the coding manual – for instance, how easy it is to use consistently. But it would not have reduced the 'subjectivity' of the coding *per se*, since the coding manual itself is the result of many decisions - decisions with which other researchers may disagree. Instead, we have chosen to focus on improving the consistency (through coding procedures), transparency (through this document) and persuasiveness (through discussion) of the coding. In other words, we have focused on ensuring that coding is consistent and transparent and that we can justify coding decisions with an argument.

Specific coding decisions for all news values (in alphabetical order)

Consonance

1. For the coding of whole units, rather than markables, three choices can be selected: ‘yes’, ‘no’, and ‘possible’.
2. An item is coded as Consonance (‘yes’), if the representation of a particular country/nation, news actor, social group, or organisation conforms to ‘obvious’, well-known, enduring stereotypes that the target audience is likely to hold.
3. An item is also coded as Consonance (‘yes’), if expectedness lexis (e.g. *typical*) is used to explicitly construct behaviour or attributes as (stereo-)typical.
4. An item is coded as ‘possible’, if the representation conforms to a potential stereotype that some members of the target audience may hold.
5. In all other cases, the item is coded as ‘no’. A ‘conservative’ approach to coding is taken so that items are coded as ‘no’ rather than ‘possible’, and as ‘possible’ rather than ‘yes’ if there is an element of doubt. References to weather events that are expected to occur in particular countries (e.g. *Super typhoon Hayan hits central Philippines*) are coded as ‘no’.
6. Since Consonance is about stereotypes, rather than expectedness, the same text can be coded as both Consonance and Unexpectedness. For example, the headline *Artist turns his dead pet into flying helicopter* establishes the news value of Unexpectedness because the described event is highly likely to be unusual for most target audience members. At the same time, it conforms to a potential stereotype that some readers may hold about ‘weird’ or ‘eccentric’ artists.
7. Since the analysis of Consonance is relatively subjective and difficult, because several different target audiences are addressed, all instances coded as ‘possible’ are documented in Table 1 to ensure transparency. In listing these, no judgment is made as to whether or not the stereotypes/preconceptions are positive, negative, have a factual basis or not. Because of the conservative approach to coding, no items were in fact analysed as clear instances of Consonance (‘yes’).²

Table 1 Text coded as Consonance: ‘possible’

News outlet	Text	Potential preconceptions/stereotypes that some members of the target audience may hold
CNN (US)	<i>5-year-old Kentucky boy fatally shoots 2-year-old sister</i>	Stereotypes about Kentucky (‘gun lovers’)
	<i>A Texas father caught a man sexually assaulting his 4-year-old daughter and punched him in the head repeatedly, killing him</i>	Stereotypes about Texans (‘take law in their own hands’)
	<i>George Zimmerman charged with felony after allegedly pointing gun at girlfriend; George Zimmerman wants to box rapper DMX in celebrity bout, but no deal yet</i>	Preconceptions about George Zimmerman (‘violent’, ‘attention-seeking’)
	<i>Arizona lawmakers pass controversial anti-gay bill</i>	Stereotypes about Arizona (‘anti-gay’, ‘socially conservative’)
	<i>A Kentucky pastor who starred in a reality show about snake-handling in church has died -- of a snakebite</i>	Stereotypes about Kentucky (‘engage in non-mainstream religious practices’)
	<i>Jim “Pee Wee” Martin acted like he’d been here before, like jumping from a plane is as easy as falling off a log.</i>	Stereotypes about soldiers/vets (‘heroic’, ‘courageous’) [the headline preceding this OP identifies Martin as D-day vet]
[continues]		

² It may be difficult for the analyst to guess what potential stereotypes members of the target audience hold if she is not a member of that target audience. The danger is that one resorts to one’s own preconceptions about the target audience (e.g. viewers of Fox News) in making assumptions about what stereotypes they may adhere to. It is for this reason that no files were coded as clear instances of Consonance (i.e. ‘yes’).

CNN (US) [continued]	<i>By most standards Matt and Paige Figi were living the American dream. They met at Colorado State University, where they shared a love of the outdoors. After getting married, the couple bought a house and planned to travel the world.</i>	Preconceptions about what it means to be 'living the American dream' [e.g. heteronormativity]
FOX (US)	<i>Arizona Sheriff Joe Arpaio puts inmates on bread and water for destroying US flag</i>	Stereotypes about Arizona ('tough love', 'patriotic')
	<i>Uwe and Hannelore Romeike came to the United States in 2008 seeking political asylum. They fled their German homeland in the face of religious persecution for homeschooling their children.</i>	Stereotypes about Germany ('rule-driven', 'intolerant')
Washington Times (US)	<i>South Carolina House passes bill making 'Obamacare' implementation a crime</i>	Stereotypes about South Carolina ('anti-Obama')
NBC (US)	<i>American student ends up trapped in giant vagina sculpture</i>	Stereotypes about students ('do stupid things', 'immature')
New York Times (US)	<i>Israelis watch bombs drop on Gaza from front-row seats</i>	Preconceptions of how ordinary Israelis behave towards Palestinians ('no compassion')
Washington Post (US)	<i>Journalists at Sochi are live-tweeting their hilarious and gross hotel experiences</i>	Stereotypes about Russia ('low quality', 'unorganised', 'exotic')
	<i>Historian believes bodies of 800 babies, long dead, are in a tank at Irish home for unwed mothers</i>	Preconceptions about Ireland and the Catholic church ('terrible treatment of unwed mothers')
	<i>In a town in western Ireland, where castle ruins pepper green landscapes [...]</i>	Preconceptions about Ireland ('romantic', 'rural')
	<i>The head of a northeast Ohio charity says that the Romney campaign last week "ramrodded their way" into the group's Youngstown soup kitchen so that GOP vice presidential candidate Paul Ryan could get his picture taken washing dishes in the dining hall.</i>	Stereotypes about politicians and/or campaign managers ('image-obsessed', 'no respect')
New York Daily News (US)	<i>Actor Paul Walker -- a self-styled adrenaline junkie best known for his role in "The Fast and the Furious" movies -- was killed Saturday in a fiery single-car wreck in Southern California.</i>	Stereotypes about 'self-styled adrenaline junkies' (engaging in risky behaviour)
Daily Mail (UK)	<i>Drug addict parents gave 23-month-old son methadone 'like Calpol' [...]</i>	Preconceptions about drug addicts ('bad parents')
	<i>Artist turns his dead pet into flying helicopter</i>	Preconceptions about artists ('crazy', 'weird')
	<i>Many animal lovers find it hard to part with their pets when they die.</i>	Preconceptions about animal lovers ('very attached to their pets')
	<i>Teenager, 19, battered dog with hammer 20 times and stabbed it through chest before leaving home to sign on at the JobCentre</i>	Preconceptions about unemployed teenagers ('behave badly')
	<i>A 19-year-old faces jail after he battered his new dog with a hammer, stabbed it in the chest and left it to die in agony.</i>	
	<i>Why three is the most stressful number of children to have</i>	Preconceptions about parents ('stressed')
Mirror (UK)	<i>Most broody mothers see having a child as a wonderful gift from God</i>	Stereotypes about mothers ('natural mothers who will cherish their child as a wonderful addition to their lives')
	<i>'Compulsory abortion for all Down's and Spina Bifida babies': UKIP candidate sparks angry backlash with website comments</i>	Preconceptions about UKIP and their candidates ('non-mainstream', 'controversial')
The Guardian (UK)	<i>Five hundred new fairytales discovered in Germany</i>	Preconceptions about Germany ('cultured')
	<i>Aldi confirms up to 100 percent horsemeat in beef products</i>	Preconceptions about Aldi ('cheap', 'low quality')
	<i>'The world stands disgraced' - Israeli shelling of school kills at least 15</i>	Preconceptions about Israeli army ('does not care about hurting civilians')

<i>News.com.au</i> (OZ)	<i>Tony Abbott lambasted on US TV Show</i>	Preconceptions about Australian PM Tony Abbott ('makes embarrassing gaffes')
<i>The Age</i> (OZ)	<i>Leonardo Da Vinci's wacky piano is heard for the first time, after 500 years</i>	Preconceptions about Leonardo da Vinci ('a crazy, quirky inventor')

Eliteness

1. For the coding of whole units, rather than markables, three choices can be selected: 'yes', 'no', and 'possible'.
2. An item is coded as 'yes', if reference is made to recognisable elite individuals (individuals such as actor *Paul Walker*, couples such as *George and Barbara Bush*, groups such as *doctors*) and organisations (such as *the federal government*), which can also be used metonymically to refer to members of the organisation (e.g. *court rules*, *South Carolina House passes*, *Obama administration pushes*). This includes different types of 'elite' people: stars/celebrities (from the entertainment industry, mainly actors, singers, stars, but including well-known directors); politicians; authority figures such as officials; royals; religious leaders; scientists/academics; writers; athletes; and managers/leaders (including business groups). It also includes different types of organisations: locally and globally important corporations (e.g. *Chick-fil-A*; *Aldi*); authorities such as a court, agency or center (e.g. *CIA*); political organisations such as a government; and NGOs/charities. The various ways in which one can construct elite status through language (e.g. role labels, proper names, status-indicating adjectives) are described in Chapter 4 of Bednarek & Caple (2017).
3. An item is coded as 'possible' where it might construct Eliteness, but the analyst has some doubts (e.g. *The world's best city is...*) or where the elite status seems weaker. This includes references to 'ordinary' soldiers/vets and police (e.g. *D-Day vet Jim 'Pee Wee' Martin*; *Police have arrested four suspects*); unnamed artists/musicians (e.g. *Drummers are natural intellectuals*); ordinary persons-turned-elite (e.g. *George Zimmerman*; *record-breaking globetrotter Graham Hughes*);³ semiotic research products without attribution (scientific journal, study, survey, conference, e.g. *Casual marijuana use linked with brain abnormalities, study finds*);⁴ and references to social media (e.g. *On Facebook clicking 'like' can help scammers*) – as long as there are no other Eliteness devices in the respective H/OP.
4. Elite countries/states/cities and events are also coded as 'possible' (e.g. *for desecrating US flags, killed in bombings at Boston Marathon*), but only those that are unequivocally recognised as elite, e.g. *UK, US, Germany, Russia* – in contrast to *Philippines, Brazil, Iraq, Nigeria, Syria, Israel, Ireland, Ukraine, Chicago, Kentucky, Essex* which are left uncoded for Eliteness, unless there are other indicators. Hence, *the Syrian government* is coded as elite because *government* refers to the political authority in Syria. While entities that can be constructed as elites include countries, nations and events (Bednarek & Caple 2017, Chapter 3), there appears to be a difference between describing the actions of elite news actors and mentioning elite nations/events, often as locations. The degree to which Eliteness is construed in the reported event seems weaker in these cases – hence the 'possible' coding.
5. In all other cases, the item is coded as 'no'. Thus, 'movements' (e.g. activists, advocates) or groups (e.g. *Jewish groups*) are not coded as elites, unless they are well-known or of high status in a given context. Neither are militants/terrorists or vague references without status-indicator (e.g. *sources, skeptics*) treated as elites. Instances of *God* (in the phrases *a wonderful gift from God*; *a story inspired by God's word*) are left uncoded. Some of the instances that are left uncoded are references to scientific findings, e.g. *Sugars found in*

³ The category of 'ordinary person-turned-elite' refers to ordinary people that have achieved considerable news fame.

⁴ Scientific studies could be considered an 'elite source', but this is more strongly the case if they are attributed to an academic institution or news actor, for example through a high-status role label (e.g. *a study by scientists from the National University of Singapore (NUS) has found that*).

tequila may protect against obesity, diabetes. These are not coded as Eliteness if no explicit reference is made to the study, institution or scientists behind the findings.

6. The *same* entity cannot be coded as both ‘elite’ (Eliteness) and ‘ordinary’ (Personalisation). For example, an instance such as *American student* is coded as Personalisation, rather than Eliteness, as this news actor is not constructed as an elite beyond perhaps belonging to an ‘elite’ nationality (USA), since students are not elite professionals. However, both Eliteness and Personalisation may be present in the same unit, as when ‘elites’ encounter the ‘ordinary’ (e.g. *Stranded Dave Matthews hitches ride with fan to show*).

Impact

1. For the coding of whole units, rather than markables, three choices can be selected: ‘yes’, ‘no’, and ‘possible’.
2. An item is coded as ‘yes’, if and only if there is reference to the consequences of a reported event *and* these consequences are either ‘significant’ (e.g. of a large scope, affecting many people) or affect the target audience directly. For instance, the following headlines are coded as Impact: *meteorites injure hundreds in central Russia*; *Boy, 8, one of 3 killed in bombings at Boston marathon, scores wounded*. Here events (meteorites, bombings) are described as having negatively affected many people. Also included are descriptions of large social media impact (e.g. *haka footage goes viral*) and described consequences for sections of the target audience, e.g. *Bad news for dads; in a landmark decision for gun rights activists; taking mummy blogs by storm*. Both actual and non-actual (future, hypothetical) explicit significant consequences are included. Impact may also be constructed through linguistic resources that assess the significance of the happening such as *Researchers might have found the Holy Grail in the war against cancer; a landmark decision*.
3. An item is coded as ‘possible’ where it might construct Impact, but the analyst has some doubts (e.g. *the pro-Russian militants who have taken over a government office in an attempt to make Ukraine part of Russia*).
4. An item is coded as ‘no’, if a reported event explicitly affects only a few people (e.g. *Parents’ fear of vaccination nearly killed their son; Father kills man sexually abusing his daughter; 5-year old Kentucky boy fatally shoots his 2-year-old sister; it’s likely the ever-popular starlet just gained more fans*) or the consequences are left unstated and only implied (*Black Rhino declared extinct; With 25 million people in its path, Super Typhoon Haiyan [...] smashed into the Philippines on Friday morning*). This captures the difference between a headline like *‘Fast & Furious’ star Paul Walker killed in car crash* (no explicit impact beyond Paul Walker) and *Fans around the world devastated by Paul Walker’s death* (invented example with explicit significant impact). Similarly, there is a difference between *Super typhoon Hayan ... hits central Philippines* (no explicit impact) and *Typhoon Haiyan death toll rises over 5,000* (explicit impact, authentic example but not part of analysed dataset). There are some borderline cases, such as *Stephen Hawking’s boycott hits Israel where it hurts: science* (affects science in Israel, but not the target audience or wider world and hence not coded as Impact).
5. An item is also coded as ‘no’, if it contains simple references to the benefit or harm to people’s health and safety as uncovered by research, without any further explicit impact constructed for the research findings (e.g. *Tequila shots may do more than lighten the mood at a party; the drink may be beneficial for your health as well*). In contrast, examples like the following are coded as Impact, since they refer to potential significant consequences: *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*.
6. Direct reader address (*you, your*, etc) is not coded as Impact, unless direct consequences for the target audience are described (e.g. *hilarious new parent test ... MIGHT just put you off [having children]*). Similarly, adjectives that describe emotional effects (e.g. *hilarious, gross, touching*) are only coded as Impact when significant effects on many are described:

Pope Francis' embrace of a severely disfigured man touches world. It was the embrace that melted hearts worldwide.

Negativity and Positivity

1. The news values of Negativity and Positivity are coded together using the label *valence*. For the coding of whole units, rather than markables, three choices can be selected: 'negative', 'positive' or 'unclear-or-none'.
2. An item is coded as 'negative' or 'positive' if the described event is highly likely to be negative/positive for most target audience members (e.g. *5-year old Kentucky boy fatally shoots his 2-year-old sister* [Neg]; *Black Rhino declared extinct* [Neg]; *Teens chase kidnapping suspect on bikes, save 5-year-old girl* [Pos]) or if it contains words with negative/positive meaning or connotations (e.g. *brag*; *best*). In other words, the aim is to take the likely attitudinal point of view of the target audience.
3. An item is also coded as 'negative' or 'positive' if detriments or benefits are described, e.g. in relation to research findings (e.g. *Casual marijuana use linked with brain abnormalities* [Neg]; *Semen is good for women's health and helps fight depression* [Pos]) or if people are labelled with negative/positive descriptors (e.g. *Drug-addict parents* [Neg]; *Drummers are natural intellectuals* [Pos]).
4. An item is coded as 'unclear-or-none' if the researcher is not certain of its valence (e.g. because she is too unfamiliar with the target audience or the target audience is likely to be divided in their attitudinal point of view). This coding is also chosen if positivity and negativity are co-present (e.g. *hilarious and gross*).

Personalisation

1. For the coding of whole units, rather than markables, three choices can be selected: 'yes', 'no', and 'possible'.
2. An item is coded as 'yes', if specific 'ordinary' people are described as involved. This includes individuals (both adults and children), pairs-of-two (e.g. *parents*), and family units (e.g. *Christian home-school family*).
3. An item is coded as 'possible' if 'ordinary' people are described as being involved, but only with generic or group references, such as *hundreds*; *knitters*, *workers*, *activists*, *residents*; *Jewish groups*; *having a child*; *the more children a mother has*. This is because we consider such generic/group references as constructing only weak Personalisation (if at all) – see Bednarek & Caple (2017, Chapter 4).
4. An item is also coded as 'possible' if 'ordinary' citizens are described as engaging in criminal behaviour (*Father kills man sexually abusing his daughter*; *drug-addict parents gave 23-month-old son methadone*; *Teenager, 19, battered dog with hammer 20 times and stabbed it through chest*; *A 19-year-old faces jail*). These differ from labels that unequivocally categorise news actors as suspects/criminals and which are coded as 'no' (see next dot point).
7. In all other cases, an item is coded as 'no'. Thus, fairytale creatures (*brave young princes and evil witches*), animals (e.g. *giraffe*, *horse*), suspects/criminals (e.g. *scammers*, *inmates*, *paedophiles*, *kidnapping suspect*), militants, terrorists etc (e.g. *rebels*), or indefinite pronouns (*someone*) and other vague references (*sources say*) are not coded as Personalisation. Further, the use of direct address (*you*) is not treated as constructing Personalisation.
8. The *same* entity cannot be coded as both 'ordinary' (Personalisation) and 'elite' (Eliteness). For instance, soldiers/vets and artists are treated as 'possible' elite professionals, and hence not coded as Personalisation. Compare the description of Eliteness above.

1. For the coding of whole units, rather than markables, three choices can be selected: ‘yes’, ‘no’, and ‘possible’.
2. An item is coded as ‘yes’, if it contains a clear reference to a place near the target audience (**same** city, state, country, e.g. *Kentucky*, *British journal*). This includes references to people from/in those regions (*American student*, *Michigan woman*, *Kentucky boy*, *state officials*, *Alaska’s senators*). Thus, a reference to Boston or Kentucky in the US news outlet CNN is coded as Proximity, as is a reference to *An American exchange student* in the NBC (US).
3. An item is coded as ‘possible’ if it makes reference to a **neighbouring or culturally close country**⁵ or contains a global reference (e.g. *worldwide*, *global*, *the world*). The latter establishes weak Proximity, as it includes the country of the target audience but is not limited to it. To give a different example, a reference to Germany in a UK newspaper (geographically close) and a reference to the US in a UK newspaper (culturally close) are both coded as ‘possible’ Proximity. They establish a weaker degree of Proximity than references to the country, nationality or culture of the target audience itself.
4. An item is also coded as ‘possible’ if it contains a **cultural reference**, whether that refers to the culture of the target audience, a neighbouring or a culturally close country. Cultural references include references to people, institutions/organisations, products, events, etc which are famous in or tied to the respective culture. Such references simply establish cultural familiarity. Examples from the corpus for the US include *George Zimmerman*, *Amy Adams*, *Obama*, *Starbucks*, *CIA*, *Glee*, *Kiss*, *MINT Shelby GT500*, *Fast and Furious*, *D-day*, *prom*, *college education*, *gun carry rights*, *Obamacare*; examples for the UK include *Richard III*, *Stephen Hawking*, *UKIP*, *Calpol*; examples for New Zealand include *haka*. To give a few examples, in a UK newspaper *Stephen Hawking’s boycott* is regarded as a cultural reference and hence coded as ‘possible’ Proximity. Similarly, *Aldi confirms up to 100 percent horsemeat in beef products* is treated as cultural reference, since Aldi is a familiar supermarket in the UK, despite its German origins. A reference to the *2014 Olympics* in a US news outlet is also considered a cultural reference because the US has hosted the Olympics in the past, the Olympics are broadcast in the US, and US athletes participate in them, although the 2014 winter olympics were hosted in Russia. It is clear that globalisation renders the issue of coding for cultural references quite complex, and furthermore that such coding is more subjective than that of geographical reference. It is for this reason that all such references are coded as ‘possible’ rather than ‘yes’. However, to code for cultural references allows us to capture the difference between *Angela Merkel* and *Barack Obama* when referenced in a US news outlet – both establish Eliteness in this case, but the latter also constructs (cultural) Proximity. In other words, the coding of cultural Proximity as ‘possible’ does not imply that cultural Proximity should **not** count as Proximity; rather it acknowledges that such references are more difficult to code.⁶ If an item contains one geographical and one cultural reference, it is coded as ‘yes’. This conforms to the general procedure number 5 outlined above (page 6).
5. In all other cases, an item is coded as ‘no’.

⁵ Cultural closeness is operationalised here as sharing the same language, so that the US, UK, Ireland, Canada, New Zealand and Australia would be conceptualised as culturally close. This is not meant to imply that language is the only basis of cultural similarity.

⁶ One could even argue that English-sounding names or nicknames constitute cultural familiarity for an English-language target audience, such as *Jim “Pee Wee” Martin* or *Graham Hughes*, even if the named persons are unknown. The assumption would be that these news actors are of a related cultural origin, even if their nationality is not stated. Conversely, one could claim that language-specific words such as *gas station* vs. *petrol station* establish (un)familiarity. However, neither English-sounding names/nicknames nor language-specific words were coded as Proximity to avoid broadening the concept too much.

Superlativeness

1. For the coding of whole units, rather than markables, three choices can be selected: 'yes', 'no', and 'possible'.
2. Not all intensifiers, quantifiers, intensified words, metaphors/similes, etc construct Superlativeness. An item is only coded as 'yes' if an aspect of the reported event is established as of high intensity or of large scope/scale. This includes but is not limited to news actors and their activities. Examples are: *an urgent request; a growing group of lawmakers; the longest-running presidential love story; a dramatic spike in; public opinion is repeatedly off the mark; ... as creepy as a Hannibal Lecter movie; an extraordinary rescue mission; most embarrassing moments; a fiery single-car wreck; a far-reaching ban; rising obesity; showed up by the thousands; something so secular; an elaborate fraud; long-lasting damage; is fighting a battle against leukemia.⁷*
3. Where *narrowly/nearly* co-occur with words denoting 'death' (e.g. *has narrowly avoided death; nearly killed*) this also seems to be used to intensify the event and is coded as 'yes'.
4. In general, the co-text and context are used to determine if a number is to be considered as relatively high. In relation to the use of numbers to report casualties or injuries, only numbers bigger than ten are coded as 'yes'. Thus, *one of 3 killed* is not coded as Superlativeness. Exceptions are where the typography, context or co-text clearly construct the number as high in scope, e.g. *Tears of a mother who lost her **FOUR** children to Chicago's gun crime epidemic* (compare Bednarek & Caple 2017, Chapter 5).
5. Concerning the coding of intensified lexis and metaphor, only those that clearly seem to construct Superlativeness are coded as 'yes'. If there is some doubt, the item is not annotated (e.g. *motorists stranded in Southern snowstorm*, where the metaphorical *stranded* and the noun *snowstorm* convey some sense of intensity/scope, but not to a large extent – compared to metaphors such as *residents battle with further snowfall* or noun phrases such as *heavy snowfall, an epic snowstorm, a blizzard*). Some other examples that are not coded as Superlativeness (i.e. are coded as 'no') include *melted hearts, tugs at the heartstrings; where castle ruins pepper green landscapes; young fighters of the disease; effort to combat rising obesity; defy the federal government; condemned a leaflet; urging.. to, sparked an ... mission; two bombs struck; hailed as heroes; stab; cheered on*. In contrast, coded as 'yes' are the intensified verbs *smash, batter, burst, ramrod; crash, rock [buildings]*; the intensified nouns *agony, invasion, [a bloody scene of] destruction; scandal; miracle; Holy Grail; the war against cancer; a disgrace to the world*; and the adjectives *irate; psyched; brilliant; perfect; startling, revolution-wracked; viral; passionate; appalling; ground-breaking*.
6. An item is coded as 'possible', when the analyst has doubts whether potential linguistic devices construct Superlativeness or not (e.g. *The most commonly used letter in the English dictionary*).
7. In all other cases, an item is coded as 'no'. This includes items with idiomatic phrases or labels ('*Greatest generation*' *veterans*) and vague comparative references that do not specify the scope of an event, e.g. *makes you feel **happier**; female-named hurricanes kill **more** than male hurricanes*. The latter clause states that both male and female-named hurricanes kill an unspecified number of people, although female-named ones kill more. Readers do not know how many people are killed by either.

Timeliness

1. For the coding of whole units, rather than markables, four choices can be selected: 'yes', 'no', 'possible' and 'not applicable'.

⁷ This coding scheme does not distinguish between higher and lower degrees of Superlativeness.

2. An item is coded as ‘yes’, if the reported event is established as happening in the near past/future (simple past/*will*, *going to* **with** temporal reference to recent past/future, i.e. no longer than ten days from the now) or being of relevance to the present (present perfect and present perfect progressive). ‘Nearness to the now’ can also be expressed through expressions such as *recent*, *amid continued*. ‘Yes’ is also chosen as coding if the event is construed as having immediacy, currency, seasonality or being ‘new’. ‘Immediacy’ is used as a cover term for a range of different instances of the present tense (simple and progressive), e.g. *DNA confirms bones are king’s*, also including (a few) more ‘timeless’ statements such as *The world’s best city is...; it’s good for their hearts; Drummers **are** natural intellectuals*.⁸ Currency or seasonality comprises references to current trends or seasonal events (e.g. a reference to snowstorms in an item published in winter, a reference to D-Day in an item published on or near D-day). Construction of events as ‘new’ is limited here to the use of *new/news* to construct ‘newness’ and to cases where *first* is used to describe that someone is the first to do something or that something has been done for the first time.
3. Both statements and questions are coded (e.g. *do you think* is coded ‘yes’, since it features present tense).
4. If an item contains several verbs and one of the verbs constructs the event as near to the now and the other does not, it is nevertheless coded as ‘yes’: e.g.: *CIA sources were denied ..., sources **say**; Teenage girl **is** left brain-damaged after smoking synthetic marijuana she bought in a gas station; Oral sex **is** good for women’s health and **makes** you feel happier, according to a study which studied the effects of semen’s ‘mood-altering chemicals’.*
5. An item is coded as ‘possible’ in doubtful or debatable cases. This includes two sub-categories of ‘newness’: ‘change’ and ‘discovery or finding’ – when these occur without any other Timeliness devices. ‘Change’ relates to instances where lexis such as *change* and *transform* is used. The construction of ‘newness’ is relatively indirect and ‘weak’ here. As the name suggests, ‘discovery or finding’ relates to reports of discoveries or findings, and includes the use of *find* to present research results or conclusions described in a separate proposition which may be left-dislocated (e.g. *Retracted autism study an elaborate fraud, British journal **finds***); the use of *find/discover* with a nominal phrase (e.g. *Scientists **find** treatment; five hundred new fairytales **discovered***), the use of *find* in court rulings (e.g. *Officials **found** guilty*), and the use of *find* with adjectives referring to bodily state (e.g. *Glee star Cory Monteith **found** dead*). These differ in the extent to which they express ‘newness’: The meaning of newness is probably strongest in instances like *five hundred new fairytales discovered*, where the direct object is expressed through a nominal phrase and refers to a concrete thing. Note that in cases where the verb is in present tense or there is some other indication of Timeliness (such as a reference to the near past/future), the item is nevertheless coded as ‘yes’ rather than ‘possible’ – in line with general procedure number 5 above (page 6).
6. An item is coded as ‘no’ if there is no explicit indication that the event is timely, although there is a tensed verb and nearness to the now may be *implied*. Thus an item containing past tense or *will/going to* **without** a specific time reference locating the event clearly in the recent past or near future is coded as ‘no’.
7. An item is coded as ‘not applicable’, if it contains modal verbs other than *will/going to* or is in imperative mood (including *let’s*) or if there is no tensed verb (e.g. *Boy, 8, one of 3 killed; Knitters wanted; Western black rhino declared extinct; Ron Pallillo dead at 64; clicking ‘like’ can help scammers; Starbucks to provide free college education*).⁹ (But if the H/OP contains a construction of Timeliness in addition – such as a reference to ‘change’ or

⁸ Differences in uses of simple present are not uncovered by this coding.

⁹ In many instances without tensed verb (excluded from coding) recency or immediacy is likely implied (e.g. *Boy, 8, one of 3 killed* → ‘is one of 3 killed’; *Ron Pallillo dead at 64* → ‘is dead’; *hilarious new parent test taking mummy bogs by storm* → ‘is taking’; *S.F. to be transformed* → ‘is to be transformed’); Some instances of modality (excluded from coding) construct timeless truth in a similar way to those with simple present tense.

‘discovery’ or a second clause with a verb in present tense, it *is* coded as ‘possible’ or ‘yes’ respectively – in line with general procedure number 5 above.)

Unexpectedness

1. For the coding of whole units, rather than markables, three choices can be selected: ‘yes’, ‘no’, and ‘possible’.
2. An item is coded as ‘yes’ if the described event is highly likely to be unexpected or unusual for most target audience members (e.g. *Knitters wanted for Penguin pullovers*), i.e. the analyst is fairly sure of her coding. The item is also coded as Unexpectedness, if lexis (e.g. *funny, wacky*) or typographic elements are used to establish or reinforce that aspects of the event are unexpected. A special sub-category of Unexpectedness is represented by descriptions of situational irony. Examples from the corpus include:
 - *Reality show snake-handling preacher dies -- of snakebite*
 - *Suicide Bomb Trainer in Iraq Accidentally Blows Up His Class*
3. An item is coded as ‘possible’ if the described event is somewhat likely to be unexpected to most target audience members or likely to be unexpected to some target audience members or the analyst is not sure of her coding. For example, that *parents’ fear of vaccinations nearly killed their son* may surprise some audience members, while others may be familiar with previous cases where a lack of vaccination resulted in dangerous consequences or might expect this to happen based on their knowledge of vaccination.
4. In all other cases, an item is coded as ‘no’

References

- Bednarek, M. (2016) Investigating evaluation and news values in news items that are shared via social media. *Corpora* 11 (2) [in press at the time of writing].
- Bednarek, M. & H. Caple (2017). *The Discourse of News Values*. New York/Oxford: Oxford University Press.
- O’Donnell, M. 2015. *UAM Corpus Tool*. Version 3.2. Available to download from: <http://www.wagsoft.com/CorpusTool/download.html>.
- Taboada, M., Carretero, M. & J. Hinnell (2014). Loving and hating the movies in English, German and Spanish. *Languages in Contrast* 14 (1): 127-161.