

Towards an algorithmic journalism assessment tool

Accounting for source diversity in local digital news

Asmaa Malik

Ryerson University

Gavin Adamson

Ryerson University

Abstract

This paper describes the motivation, progress and early results in the development of an algorithmic tool for evaluating journalism called JeRI, or the journalism representation index. Using named entity extraction and classification plus machine learning techniques, the project is aimed at building software that identifies, quantifies, and categorizes the sources quoted or attributed in a news story. The goal is to provide a score for any textual piece of news journalism that would be available to all stakeholders in the news – the newsrooms themselves, readers, and social scientists. The software redeploys a coding rule for sources from Poindexter, Smith, and Heider (2003) that underlines how local journalism prefers sources that support status quo and under-represents citizens and activists. Using a corpus of articles about police carding obtained online from the Toronto Star’s website, the paper describes the process of training a classifier to automatically gather basic facts about each article – about the number, frequency, and the types of sources that the text refers to. The study tests whether JeRI can make the simple judgments about sourcing that humans can. The tool’s determinations about sources are at the heart of an index that would score any article based against industry averages. JeRI is described in the context of framing analysis but also in the context of the state of the digital, algorithm-driven journalism industry.

Keywords

local digital news, computational journalism, algorithms and journalism, race and policing, natural language processing

Introduction

Newsrooms, the audience, and scholars are interested in the quality of journalism and they spend a great deal of time and energy making judgments about it. Internally, news editors ask reporters to adhere to editorial standards of fairness and balance (for examples see “Ethical journalism...,” 2018; “Journalism code of ethics...,” 2018) that are supported and upheld by journalism schools (for example see Perkins, 2016). Media watchdogs and citizens hold news organizations accountable for their representations of authorities, institutions, and marginalized groups (Ho, 2018; Sylvester, 2016). Many studies across a variety of fields in the social sciences describe that the news media tend to support the status quo. Hardly a social sciences conference happens today that doesn’t include discussion of a framing study that critically analyzes how journalists select and interpret facts for their stories (Entman, 2010; McNair, 2009). And of course, social platforms give readers opportunities to express their reactions with ‘likes,’ emojis, and hearts, and they share their thoughts about journalism via commenting features.

As corporate communications, police and governments increasingly rely on the Internet to distribute their preferred versions of stories (Cheung & Wong, 2016) and as journalism suffers from a lack of resources to challenge those views (Van Leuven, Deprez, & Raeymaeckers, 2014) stakeholders’ concerns about the quality of the news is especially cogent and timely. While there is no shortage of opinion about the issue, at the same time there is also not one consistent benchmark by which journalism can be judged. Framing research boasts a rich field of theory with some of it pointing to sourcing, but even its proponents suggest many studies are subjective, the methods opaque, and the results neither repeatable or valid (Matthes, 2009; Van Gorp, 2009).

The current paper describes a project aimed to construct one such benchmark. It postulates that journalism sources can be at the focus of a scoring system, called JeRI, or the journalism representation index. *To whom do journalists speak? Who do they quote? From whose point of view are stories told, and whose voices get the most prominence?* JeRI aims to apply a transparent, computational approach to scoring journalism that takes a cue from the journalism process (i.e. talking to people). It takes a cue from social sciences research while keeping the act of journalism in mind to

underline that questions above (in italics) are critical. As news is being produced and distributed algorithmically by social platforms, this approach posits that there is a place for equally immediate and transparent judgment about the choices.

Researchers have taken computational approaches to analyzing journalism texts in the past (Crowston, 2012) but results are limited by the ability of software to understand the complexity of language. While this study similarly involves training software to recognize patterns it simplifies the process by focusing on the sources in the text, which is significantly less fraught. The method described in the present study involves (1) obtaining a digital corpus of stories about police carding and racial profiling appearing in Toronto online media; (2) extracting the sources from the text; (3) identifying, counting, and coding the sources by humans as per Poindexter, Smith, and Heider (2003); and (4) training a classifier to assign categories of the sources that meet human coder reliability.

The preliminary results indicate that named-entity (in this case, source) recognition is advanced enough to be used to recognize journalism sources in text and for social scientific analysis. Early results show that together, entity extractors can indeed identify and categorize sources. Still, many ambiguities in context suggest that success of the method may also be limited by the extent to which the human coders can agree about categorizing sources.

Sources and frames

From a theoretical and methodological point of view, JeRI applies two mass communications theories to quantifying, weighing, and accounting for sources in journalism. Each of them involves consideration of sourcing. Sourcing is one of the elemental journalistic practices and has also been the subject of analysis for a variety of mass media researchers (Franklin & Carlson, 2010). Rauhala et al. (2011, p. 96) defined sources as: “those who supply the content for stories, provide context to a narrative, offer opinions, and give witness accounts.” These are “often officials and experts connected to society’s central institutions” (Berkowitz, 2009, p. 113). Journalism texts have regularly noted that the first source mentioned in a story exerts a strong influence on the reader (Friend, Challenger, & McAdams, 2003; Russial, 2004).

Researchers from a variety of disciplines and theoretical backgrounds take a critical approach to sources. In his theory describing how journalists speak with both government and opposition sources, Bennett described the “indexing hypothesis” (1990) in the context of press and federal government relations in the U.S. He describes

that the news media “index” the sources in their reporting and in editorials only according to the range of views narrowed in the government debate about a given topic. Dissenting opinions are only written about when they ultimately reflect news in their beat through the official sources, while political views that diverge greatly from the norm are very infrequently expressed, according to his theory. In other words, government sources get the chance to contextualize even opposing points of view, and “this tends to reflect the perspectives promulgated by those whom journalists perceive to have the most power to influence a situation” (Lawrence, 2010, p. 269).

Indexing accounts for reporters’ sourcing choices in a variety of political beats, particularly in the context of foreign political reporting (Bennett, Lawrence, & Livingston, 2006; Entman, 2010; Kennis, 2009). One study also indicated that official sources could control the timing of important news and, in turn, that timeliness (i.e. the need for journalists to report the latest information) helped explain indexing routines (Althaus et al., 1996). Dominant, official sources also meet journalistic demand for immediate response in a shortened, live news cycle, according to the hypothesis (Livingston & Bennett, 2003). The indexing pattern is less obvious in municipal government reporting, but domestic reporting often indexes similarly when law enforcement is concerned (Lawrence, 2010). The indexing theory underlines the prominence of repeated, timely, status-quo political messaging in political journalism. The theory also takes an empirical approach to analyzing journalism texts, using frequency of data as a variable. For these reasons it plays a key theoretical role in the evaluative tool we are developing.

Bias, frames, and sources

Sourcing is also often a consideration in theoretical discussions about how frames are built in journalism stories (D’Angelo & Kuypers, 2010). Generally, frames work as mental heuristics that allow individuals to process information. Framing methodology involves making a judgment about journalism texts and productions about their bias, in the most general sense of the word (Entman, 2010). To “frame” is most very often defined in the literature as the act of selection by the news media, “in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described” (Entman, 1993, p. 53). Researchers make different assumptions about how they are manifested culturally, within the minds of the public, within journalistic norms, or in combination with other actors such as politicians, lobbyists, and other expert sources (D’Angelo, 2002; Entman, Matthes, & Pellicano, 2010; Matthes & Kohring, 2008; Scheufele & Scheufele, 2010). In one clear example, a researcher showed that local newspaper coverage about

the petrochemical industry in Spain tended to be framed in the news as a local economic driver, while environmental frames, for example, were secondary (Castello, 2010).

Theoretical meta-research about framing itself has shown that it can be fraught. For this reason we do not aim to – nor could we, as we describe below – translate framing theory into a computational judgment. Theorists in this field have underlined that the validity and reliability of framing studies are a concern in this realm of research because the manner by which frames are identified is not always clear (Gamson, 1992; Miller, 1997; Tankard, 2001). Researcher subjectivity is a concern (Matthes, 2009; Van Gorp, 2009). At the same time, researchers have called for more, “theorizing and theory-building research on source selection, source use, and source impact” (Strömbäck et al., 2013, p. 14; Carragee & Roefs, 2004).

Crucially, regardless of the variety of approaches and considerable disagreement in this “fractured” research paradigm (Entman, 1993, p. 51), sources are almost always a central consideration in the production of frames (Borah, 2011, p. 246; Gans, 1979; Scheufele, 2000; Van Gorp, 2009). Dovetailing with Bennett’s hypothesis, framing research describes an over-reliance on “official” sources that support existing power structures (Dimitrova & Strömbäck, 2012; Bae & Lee, 2012; Fahmy & Eakin, 2013; Houston, Pfefferbaum, & Rosenholtz, 2012; Lee & Basnyat, 2013; Mahfouz, 2013; Strömbäck et al., 2013). Choices that reporters make about sources can be considered decision-making bias (Entman, 2010). The news media do not merely mirror the frames of their sources but do work that exists somewhere between frame sending and frame setting. In producing a report or an analysis a journalist may reflect her individual frame, she may almost purely reflect that of a source, or she may produce an “interpretative” (Brüggemann, 2014, p. 64) report that includes a mixture of her sources’ frames. As sourcing is described as perhaps the seminal journalism act and because researchers have underlined its importance in the critical literature, our journalism assessment tool makes the evaluation of sources the very point of its processing. The concentration on reporterly choices – about who they speak to, and how often they quote sources, is a key concept from both framing and indexing theory with which JeRI is concerned.

Transparency in sourcing as an algorithmic process

JeRI’s automated approach to transparency acknowledges that journalistic authority has become “built on openness rather than the old ‘we write, you read’ dogma,” and now news organizations are expected to explain their processes and their decisions

(Karlsson, 2011, p. 280). *The Guardian's* Datablog, for example, offers readers the opportunity to download data sets featured in their major investigative projects and replicate their journalists' work. Several news organizations including National Public Radio and *Wired* magazine fact-checked the 2016 U.S. presidential debates, comparing candidates' statements on live television with comments and promises they had made earlier on the campaign trail. Digitalization has come with the move toward greater transparency online (Deuze, 2005; Karlsson, 2011), yet readers are offered only limited exposure to journalistic processes. Journalists must do more to disclose their methods of newsgathering so readers can reach their own conclusions based on the information news organizations publish (Kovach & Rosenstiel, 2014).

Algorithms have transformed the way journalists and audiences interact with news. Media organizations distribute user-targeted information in as many different ways as there are digital devices, struggling with increasingly limited resources to repurpose content as much as possible, as quickly as possible (Karlsson, 2011; Karlsson & Strömbäck, 2010). Social networks deliver news that is tailored to fit readers' ideologies and interests, as well as reinforce partisan allegiances (Bakshy, Messing, & Adamic, 2015). "After disrupting our distribution channels, digitalization reaches out for our production," writes Mercedes Bunz in *The Silent Revolution* (2014, p. 5). "Automated journalism" refers to news bots created and employed by universities and wire services that use natural language generation (Dörr, 2016) to write instant sports and business stories, bypassing human intervention to publish nuts-and-bolts articles that share the relevant facts of a college basketball game or a company's earnings reports.

Now that news organizations and their readers have become more familiar, if not yet completely comfortable, with the digitalization of news distribution as well as production, the next step is inevitable. "The technology has the potential to disrupt everything that implies coherent information, and this means: our expertise" (Bunz, 2014, p. 5). Algorithms already drive much of the content readers view and consume online, and through functions such as social networks and web search they have also become an important part of the newsgathering process. These shifts in distribution and production pave the way for the creation of algorithmically-powered tools, such as JeRI, that give readers and journalists the opportunity to analyze and evaluate the news they read.

With speedy-delivery web and mobile technologies such as Facebook Instant and Google's Accelerated Mobile Pages Project, which aspires to create articles that load "near instantaneously" (Dawson, 2016), readers have come to expect immediate

access to the news they want to read. The pace of digital news, which relies on constant updates and near-instant clarifications of incorrect information, has driven news organizations and their audiences to an openness in communication (Plaisance, 2007) leading to a greater transparency in journalism (Deuze, 2005; Karlsson, 2011; Karlsson & Strömbäck, 2010). Research into sources “sheds light on the quality of information, the diversity of perspectives and the interpretive framework on offer” to readers (O’Neill & O’Connor, 2008, p. 488) and as news organizations struggle with limited resources and fewer journalists to produce original reporting (Maat & de Jong, 2012), evaluating “who gets ‘on’ or ‘in’ the news” (Cottle, 2000, p. 427) offers a critical perspective into the reporting process. Mainstream sources – especially “parajournalistic” ones such as public relations firms, media officers, and spin doctors (Schudson, 2003, p. 3) – are particularly favoured compared to “ordinary citizens and NGOs.” The trend toward “desk journalism” has led to a dependence on the telephone for quick interviews and story ideas (Davies, 2008), leading journalists to favour sources on their speed dial. JeRI’s goal is to make transparent journalism source choices, but various computational challenges need to be considered first.

Computational approaches to analyzing journalism texts

Researchers have described how computer science can help in academic research using “algorithmic or computational solutions to generate patterns and inferences from data” (Shah, Cappella, & Neuman, 2015, p. 7). Textual analysis normally comes with laborious, time-consuming manual coding (McFarlane, 2011). Computational textual analysis offers the promise of quickly reducing the amount of analysis time by a numeric factor (Crowston, 2012) once a machine has been programmed to do the work. Relating this approach to framing analysis, a meta-study by Matthes (2009) showed that 8 of 131 articles about framing put into practice computational analysis. The process replaces some of the human-coded judgments and classification of texts with statistical analysis, applying probability equations to identify phrases in large sets of text (McFarlane, 2011). Miller et al. (1997) described “frame mapping,” which assumes frames are manifested in groups of words that tend to occur together. They applied clustering techniques to texts. Stitched together along with other mathematical rules, these algorithmic calculators make classification systems. This process begins when researchers identify the universe of words that mark the presence of a frame. For example, Kellstedt (2003) programmed a machine to track words such as “individualism” and “egalitarianism” as part the proxies for the egalitarian frame in his study about the attitudes toward government efforts to promote racial equality. After integration of the dictionary, the software analyzed 6,500 articles.

The literature also underlines challenges related to algorithmic social science as anticipated by Krippendorff who described the impossibility of programming a machine to understand subtleties of language contained in metaphor and tone (Zamith & Lewis, 2015). For example, one algorithmic approach involved testing a classification system to identify health frames from journalism texts. The method ultimately reflected low agreement scores between human coders at the coding stage (McFarlane, 2011). By contrast, Zamith and Lewis note that computational methods are best applied to texts whose variables are easily identified in texts that can be easily parsed (2015, p. 313). Researchers have argued that algorithmic approaches yield satisfactory results when restricted to structural features of text (Conway, 2006; Sjøvaag & Stavelin, 2012). Our approach sidesteps the challenges described above by limiting the computational textual analysis to identifying, quantifying, and categorizing nouns, or named entities, in computer science terms. To our knowledge, named entity extraction algorithms, which are constantly refined and tested annually (Marrero, Sánchez Cuadrado, Morato Lara, & Andreadakis, 2009), have not been used at the core of a social science approach such as the one we describe below. Algorithms such as these are often made available on open-source platforms and applied to digital texts attainable by scraping websites (Zamith & Lewis, 2015). The aim of the experimental software is to identify, categorize, and ultimately score the variety of sources in a corpus of journalism texts that makes meaningful distinctions between stories that prioritize police and politicians and those that seek source diversity and depth.

Race, police, and local journalism

Although this present study does not aim to identify people of colour in journalism content, race is part of the discussion because journalism sourcing is frequently at the centre of studies about how race and power are reflected in the news media. A basic assumption in this background literature, including the framing approaches described above, is that journalism is socially constructed and that its product, the news, is defined by the context within which it is gathered and developed (Spratt et al., 2007). Studies are frequently critical about how journalists source their stories in a manner that prefers authority because the practice tends to “legitimate or even reify the power structure of society (Berkowitz, 2009; Manning, 2001; Sigal, 1973; Soloski, 1989). News has been shown to under-represent private citizens in favour of those with more authority and power, especially when the private citizens are people of colour (O’Neill & O’Connor, 2008; Poindexter, Smith, & Heider, 2003). Black people were less likely to be experts, company spokespeople, or government sources in broadcast news than white people and they are more likely than white people to be ordinary citizens in reports (Owens, 2008). Owens describes the “incognizant racism” that is identified by

such studies: “... journalists are following the same routine, unconsciously interviewing the same types of sources, and therefore producing the same kind of coverage day after day” (2008, p. 356). Anticipating the methodology section below, just for the moment, the software is centrally concerned with the routine textual repetition and prominence of the types of sources – ‘authorities’ such as police, politicians, for example – that favour status quo interpretations or reactions to political movements. The software takes up that challenge while also taking on calls for more transparency about the decisions that journalists make.

To test our software, we’ve gathered a corpus about police racial profiling and carding, a now-discontinued but long-term practice under which Toronto Police Services kept a confidential database of personal information about suspected criminals (Gillis & Rankin, 2015). Young black men were disproportionately targeted under the policy, which was heavily criticized for perpetuating racial profiling in the force (Rankin & Winsa, 2012). There is reason to believe the sourcing in the corpus will be heavily tilted towards the police as per Poindexter, Smith, and Heider (2003). As declining advertising revenues and shrinking audiences continue to constrain newsrooms and their human resources, they are increasingly more open to story pitches from public relations firms (Van Leuven, Deprez, & Raeymaeckers, 2014) and in many cases, government institutions and police forces have taken a page from the sophisticated PR strategies employed by technology firms and startups. In Canada, Hong Kong, and the United Kingdom, police forces have replaced “analog police communications” (e.g. radio scanners) that could be easily accessed by reporters with “digital police communications,” which force them to rely on information distributed by the police (Cheung & Wong, 2016). In Toronto, for example, citing security issues, police have switched to encrypted digital radio technologies and now direct journalists to their Twitter feed for live, curated information on police- and crime-related news events. “We will not be able to duplicate everything that people got by eavesdropping on the radios – that’s simply not possible,” said a Toronto Police Services spokesperson (Otis, 2015).

In their survey of 2,900 news stories published in four West Yorkshire papers, O’Neill and O’Connor found police services have have “highly developed media relations.” Instead of being able to directly interview detectives about a particular investigation, reporters must speak to media relations professionals, losing out on essential background details (2008). Shrinking staffs have led to the dissolution of the beats-reporting system in newsrooms across North America. Many are “depending more than ever on a staff of generalists to cover the news” (Ward, 2015). With the disappearance of specialized journalists who cultivate relationships with their sources

and the pressures of immediacy brought on by the digitalization of news, we have seen the proliferation of the one- or two-source story (O'Neill & O'Connor, 2008). In many cases, non-beat-reporters' interactions with police officers are limited to questions at news conferences. In its aptly named Corporate Communications department, Toronto police force employs Media Relations Officers as well as Social Media Relations Officers to respond to Twitter and Facebook queries (Toronto Police Services website, 2016). Related personnel also post press releases on selective investigations and routinely livestream and archive news conferences on their YouTube channels. Pressed for time and resources, general assignment reporters on digital desks often report on or tweet live police news conferences, essentially transcribing what the spokesperson says. This can lead to the perpetuation of police jargon and more importantly, reinforcement of the police force's preferred narrative. Research about news sources in Toronto-based police news stories, found that 14 of 26 stories, or 53 per cent, relied upon the police as the single source, or the stories focused on police-driven news such as a press conference or police action (Lindgren, 2015). The remaining 12 stories involved police spokespeople, but they were also populated by quotes from witnesses, family members, and neighbours. These stories tended to be investigations, crime trends, or criminal court coverage.

Toronto Star columnist Desmond Cole, an outspoken activist against police carding and racial profiling, criticized the newspaper and other local news outlets for repeating a police spokesperson's description of a young, white bank robber as the "Preppy Bandit" in headlines and social media posts. Contrasting the man's portrayal by police to that given of racialized suspects, Cole said "we reserve our worst assumptions and judgments about crime for the poor, the racialized, for trans and gender nonconforming people, for the uneducated" (Cole, 2016). His concerns about police language were further reinforced several days later when local news organizations readily repeated Toronto Police Service's description of the arrest of 16 racialized men in relation to local robberies, calling them a group of "pathetic parasites." Several news sites used the phrase in quotes but with no attribution specifically to police's use of the term (Doucette, 2016; Fox, 2016; McGillivray, 2016).

Methodology: What we have achieved so far and what we aim to do

It was important from the outset to establish an open digital-information retrieval process, because we envision JeRI operating automatically, i.e. analyzing the article as it is opened by the reader. For the pilot project, to create a digital corpus of 572 articles related to carding and racial profiling, we used a Python and BeautifulSoup script to scrape articles from the *Toronto Star* using the search term "carding" between January

1, 2015 and December 31, 2015. We identified open-source natural language processing tools that could identify “named entities” (or personal pronouns) from texts. AlchemyAPI and OpenCalais, two named-entity extractors, were the best candidates as per Rizzy and Tronc’s named entity disambiguation test website.

At the beginning of the study, human coders produced a so-called “gold standard” as per McFarlane (2011) by reading the text and annotating all sources according to Poindexter, Smith, and Heider (2003) for journalism texts. Coders identified sources in texts and categorized them into: (1) unaffiliated private citizens, including witnesses, consumers, students and voters (UNA); (2) authorities, including government officials, politicians and police (AUT); (3) experts, including academics, physicians and lawyers (EXP); (4) organizations, including spokespersons for businesses, nonprofits and advocacy groups (ORG); (5) celebrities, including sports figures and artists (CEL); and (6) media figures and news organizations (MED). We repeated annotation until high intercoder reliability was achieved.

In the next phase, we tokenized all the text in the corpus with an open-source Natural Language Toolkit (NLTK). Tokens are the words and word phrases in the text as described in computer science terms. For example, “Coun. Smith” is a token that would be recognized by one of the entity extractors above, and it becomes a token that can be annotated by the software. Human coders are progressing with annotating the tokens in the corpus using their gold standard. For example, the phrase “Coun. Smith” above would be identified as an authority (AUT), which includes politicians or candidates. In the next phase of our work, we will use the annotated gold standard to train a custom CRF (Conditional Random Field) classifier and Support Vector Machine (SVM) framework as described by previous researchers to iteratively test and refine the machine (Bratus, Rumshisky, Khrabrov, Magar, & Thompson, 2011; Ekbal, Saha, & Sikdar, 2013). In this process, given iterative feedback, the software learns to make more accurate statistical predictions about language. Once the tokens are annotated with sources from all of the articles in the corpus, the aim is for JeRI to identify and categorize sources with a high enough degree of accuracy with which a scoring index can be built mathematically.

A final part of the design involves building a mathematical index by which the tool will ultimately score any article, based on the depth and thoroughness of its sourcing. From our corpus of articles about carding in Toronto, the machine will be able to identify (1) the number of sourcing incidents per category; (2) average source order for the categories; and (3) the weight of the source in the text based on words quoted or paraphrased. Any single article in the corpus (or a subsequent one) can be

compared to the average and assigned an index score based on a coefficient calculated in relation to the average score for the entire corpus. For example, subject to real analysis, a story about racial profiling that contains a single political source’s opinion about carding would be scored low score, reflecting that a single source is inadequate in several ways. First, it reflects only one person’s perspective; second, as a political leader that person’s viewpoint likely does not diverge from the political mainstream and supports the status quo; and third, we hear frequently from this category of voice about racial profiling, compared to less powerful sources, such as political activists and citizens. The design and calculations in the index will be as iterative and as transparent as the algorithm itself. Crucially, the numeric judgment could be offered instantaneously and annotated with explanatory theory in the user interface of an application or a simple widget.

Table 1: *Descriptive results of human coding and JeRI tokenization (reflecting first 50 stories coded)*

<i>Source by category</i>	Coders	JeRI: Entered
Number of sources/total	395	489
AUT (Authority)	227	44
EXP (Expert)	63	23
UNA (Unaffiliated)	63	17
ORG (Organization)	4	1
MED (Media)	38	1
Skip*	n/a	403

**Skip indicates instances in which JeRI misidentified sources and then was trained to categorize them correctly. JeRI results are expected improve as training progresses.*

Table 2: *Descriptive results of human coding of Toronto Star story, “Anti-carding coalition urges a bold rewrite of Ontario’s rules for ‘street checks’” (December 5, 2015)*

Source	Frequency	Position	Attribution words	Lines of attribution	Headline	Subhead	
Coalition	1	1	Says, Calls, Say, Reads, Continues the release	8	1	1	
Howard Morton	1	2	Says	1	0	0	
Alok Mukherjee	1	3	Said	1	0	0	
Police leaders/unions	1	4	Said	1	0	0	
Rights groups	1	5	Hailed, Write, Say	3	0	0	
Sukanya Pillay	1	6	"Voices"	1	0	0	
Renu Mandhane	1	7	"Voices"	1	0	0	
Knia Singh	1	8	"Voices"	1	0	0	
Gordon Cressy	1	9	"Voices"	1	0	0	
Cutty Duncan & Shadya Yasin	1	10	"Voices"	1	0	0	
Ranjit Khatkur	1	11	"Voices"	1	0	0	

Descriptive results

At the time of publication, human coders have categorized sources in all of the stories in the corpus and have begun annotating the tokens in JeRI itself. Preliminary results reflect previous studies that show how tilted the journalistic enterprise is away from private citizens and political activists. Of the 395 sources identified by human coders in the first 50 stories (in chronological order), 227 (or 57 per cent) of them have been categorized as authorities (AUT), e.g. government officials, politicians, and police.

In the initial tokenization phase, JeRI over-identified the number of sources that appeared in the first 50 stories. The rate at which the software can categorize the sources is low because the software is identifying both the personal names and their titles as separate sources. For example, in the case of “Toronto Police Services spokesperson Const. Victor Kwong,” JeRI yielded three distinct listings for “Toronto

Police Services,” “Const.” and “Victor Kwong.” Once the challenge is corrected, the rate at which the software will correctly identify the source categories will be much higher. JeRI’s accuracy is also expected to increase as the human coders continue to identify sources as part of the tokenization process. When it came to correctly categorizing sources in the first 50 stories, JeRI was most successful at identifying experts (EXP), with 36.5 per cent accuracy, and unaffiliated individuals (UNA), with 27 per cent accuracy. Overall, the software correctly identified 21.8 per cent of the sources categorized by human coders. However, it identified just 1 example of 38 media sources (MED), 2.6 per cent, in the corpus.

The second table shows one sample of the descriptive results of human coding tasks. For example from the *Toronto Star* story, “Anti-carding coalition urges a bold rewrite of Ontario’s rules for ‘street checks,’” published on December 5, 2015, coders were able to identify if a source was named first, if the source was named in the headline or subheadline, and how many lines of text (ie, quotes, paraphrases) were attributed to the source. The “coalition” was named once, as a first source in the story. It was also named in the headline and subheadline, and given 8 lines of attribution using several different attribution words. “Howard Morton” was named once in the article, in the second source position, and so on. In subsequent iterations, JeRI can be trained to make similar observations about each story. This information is essential to the calculation of the JeRI index score, as outlined above.

Discussion

The low scores for some categories of source recognition are to be expected. The tokenization process will build a library of tokens related to the categories in the corpus. As more tokens are added to the category libraries, the tokenization process builds toward a higher rate of accuracy as the software is trained with more news articles. For example, “CBC” was unrecognized by the software as a source but is expected to be recognized on a subsequent test. For more complicated sources, the software will make “guesses” on the source categorization based on the CRF (Conditional Random Field) classifier and Support Vector Machine (SVM) as per the methodology described above.

Significant challenges may prevent the software from reaching a 100 per cent coding rate. The human coders, for example, note the difficulty of identifying sources who may appear with expert testimonial in some articles while appearing as citizens in others. Another issue that arises with the categorization of sources by JeRI is that it cannot differentiate in instances where a single person can be categorized differently

depending on the context. For example, a former chair of the Toronto Police Services Board was classified by the human coders as an authority (AUT) when he was identified as such but in cases where he was referred to as an advocate against police carding, he was categorized as an expert (EXP). In another instance, a law student who was a victim of police carding was identified as unaffiliated (UNA) in stories about the case but as an expert (EXP) when he was being interviewed as an activist against the practice. This kind of ambiguity requires extensive refinement in future iterations.

In some instances, JeRI identified entities mentioned, but not specifically attributed, in the text. For example, in a story about the Toronto police chief wanting to retain data obtained from carding, a lawyer is referenced as having filed a challenge against the information being kept. However, the lawyer is not quoted or paraphrased in the story. To prevent JeRI from counting mentions as sources, we can consider refining the software to search for named entities within a certain proximity to commonly-used journalistic attribution phrases preferred by news organizations including Canadian Press, such as “said,” “explained,” and “asked.” Together, we do not know what percentage these coding challenges will represent for the software but through training iterations the aim will be to resolve them to an acceptable level approaching 80 per cent as per Krippendorff’s alpha.

The final phase of the project, which will we undertake once we have reached an acceptable rate of accuracy through the tokenization process, will be determining the JeRI index score. Using the methodology outlined above, sources are weighted for *power* (i.e. authority figures, experts, unaffiliated), *placement* (where they are mentioned in the story hierarchy) and *frequency* (how many times names and categories appear in the article). These types of considerations are included in Table 2 above. We consider the counting of sources, the identification of their attribution position, and whether the sources are named in the headline or subheadline to be less challenging tasks for software. In any case, the index score will be a transparent statistical calculation combining coefficients for the weights above.

Conclusion

The desired outcome of this project is to provide a report card, effectively, on journalism, with a focus on sourcing practice. Understanding sourcing tendencies in coverage might inspire new approaches to certain subjects. A newsroom may be able to apply some findings to their future work. Such feedback could help establish a framework that would introduce some level of accountability to newsrooms and news-gathering. Media watchdog organizations can also use the data gathered from

the analysis to evaluate coverage on issues critical to their communities. In the case of local news organizations, reports on their current sourcing practices on an issue such as race and policing may lead to the inclusion of more diverse perspectives on the topic. Additionally, a deeper reporting strategy that seeks to avoid the “usual suspects” and favours unaffiliated citizens in news stories of major public interest. As news organizations struggle to stay competitive across platforms, devices and networks, it becomes increasingly more important that the news they produce reflects the new audiences they hope to draw.

The opportunity to generate a near-instant JeRI index score when the reader calls up a story, one that shows the reader how much or how little the news article favours the status quo as represented by sources at the top of the power hierarchy, is very much in line with the idea of empowering audiences to reach their own conclusions based on algorithmic evidence. Computational journalism tools are designed with the view that the audience is “autonomous and creative enough to perform their own searches of data” (Gynnild, 2014, p. 725). JeRI starts with assumption that news stories favour more socially powerful sources, to that extent the work of developing the application is akin to “reverse-engineering,” which Diakopoulos (2015) calls “the process of articulating the specifications of a system through rigorous examination drawing on domain knowledge, observation and deduction to unearth a model of how that system works.” While our pilot project will be looking at race and policing, subsequent areas of focus include local and national coverage of mental health, Indigenous issues, and climate change.

Our approach helps journalists keep up with the fast pace at which they report and produce news, and helps audiences match the speed at which they receive and consume news. In our current era of concerns about “fake news,” aggregation, and republished news on websites, as well as social media and newsletters, JeRI is a valuable accountability tool that can be used on its own or in addition to verification and contextual tools for a more complete picture of the people and processes behind the news we so rapidly produce and consume.

References

- Althaus, S. L., Edy, J. A., Entman, R. M., & Phalen, P. (1996). Revising the indexing hypothesis: Officials, media, and the Libya crisis. *Political Communication*, 13 (4), 407–21. <https://doi.org/10.1080/10584609.1996.9963128>
- Bae, Y., & Lee, H. (2012). Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular Twitterers. *Journal of the American Society for Information Science and Technology*, 63 (12), 2521–35. <https://doi.org/10.1002/asi.22768>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348 (6239), 1130–32. <http://doi.org/10.1126/science.aaa1160>
- Bennett, W. L. (1990). Toward a theory of press–state relations in the United States. *Journal of Communication*, 40 (2), 103–27. <https://doi.org/10.1111/j.1460-2466.1990.tb02265.x>
- Bennett, W. L., Lawrence, R. G., & Livingston, S. (2006). None dare call it torture: Indexing and the limits of press independence in the Abu Ghraib scandal. *Journal of Communication*, 56 (3), 467–85. <https://doi.org/10.1111/j.1460-2466.2006.00296.x>
- Berkowitz, D. (2009). Reporters and their sources. In Wahl-Jorgensen, T. & Hanitzsch, K. (Eds.), *The handbook of journalism studies* (pp. 101–115). London, England: Routledge.
- Borah, P. (2011). Conceptual issues in framing theory: A systematic examination of a decade’s literature. *Journal of Communication*, 61 (2), 246–63. <https://doi.org/10.1111/j.1460-2466.2011.01539.x>
- Bratus, S., Rumshisky, A., Khrabrov, A., Magar, R., & Thompson, P. (2011). Domain-specific entity extraction from noisy, unstructured data using ontology-guided search. *International Journal on Document Analysis and Recognition*, 14 (2), 201–11. <https://doi.org/10.1007/s10032-011-0149-5>

- Brüggemann, M. (2014). Between frame setting and frame sending: How journalists contribute to news frames. *Communication Theory*, 24 (1), 61–82.
<https://doi.org/10.1111/comt.12027>
- Bunz, M. (2014). *The silent revolution: How digitalization transforms knowledge, work, journalism and politics without making too much noise*. London, England: Palgrave Macmillan.
- Carragee, K. M., & Roefs, W. (2004). The neglect of power in recent framing. *Journal of Communication*, 54 (2), 214–233. <https://doi.org/10.1111/j.1460-2466.2004.tb02625.x>
- Castello, E. (2010). Framing news on risk industries: Local journalism and conditioning factors. *Journalism*, 11 (4), 463–80. <https://doi.org/10.1177/1464884910367592>
- Cheung, M. M. F., & Wong, T. C. (2016). News information censorship and changing gatekeeping roles: Non-routine news coverage and news routines in the context of police digital communications in Hong Kong. *Journalism & Mass Communication Quarterly*, 93 (4), 1091–114.
<https://doi.org/10.1177/1077699016628818>
- Cole, D. (December 21, 2016). The power of being ‘preppy.’ *Toronto Star*. Retrieved from
<https://www.thestar.com/opinion/commentary/2016/12/21/the-power-of-being-preppy-cole.html>
- Conway, M. (2006). The subjective precision of computers: A methodological comparison with human coding in content analysis. *Journalism & Mass Communication Quarterly*, 83 (1), 186–200.
<https://doi.org/10.1177/107769900608300112>
- Cottle, S. (2000). Rethinking news access. *Journalism Studies*, 1 (3), 427–48.
<https://doi.org/10.1080/14616700050081768>
- Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, 15 (6), 523–43. <https://doi.org/10.1080/13645579.2011.625764>

- D'Angelo, P. (2002). News framing as a multiparadigmatic research program: A response to Entman. *Journal of Communication*, 52 (4), 870–88.
<https://doi.org/10.1111/j.1460-2466.2002.tb02578.x>
- D'Angelo, P. & Kuypers, J. A. (2010) *Doing news framing analysis: Empirical, theoretical, and normative perspectives* (1st ed.), New York, NY: Routledge.
- Datablog. (2016). *The Guardian*. Retrieved from: <https://www.theguardian.com/data>
- Davies, N. (2008). *Flat Earth news*. London, England: Chatto and Windus.
- Dawson, J. (2016, October 20). Platforms like Facebook's Instant Articles and Google AMP are making it harder, not easier, to publish to the web. *Recode*. Retrieved from:
<http://www.recode.net/2016/10/20/13318746/online-publishing-platform-amp-facebook-instant-articles-apple-news>
- Deuze, M. (2005). What is journalism?: Professional identity and ideology of journalists reconsidered. *Journalism*, 6 (4), 442–64. <http://doi.org/10.1177/1464884905056815>
- Diakopoulos, N. (2015). Algorithmic accountability. *Digital Journalism*, 811 (February 2015), 1–18. <http://doi.org/10.1080/21670811.2014.976411>
- Dimitrova, D. V., & Strömbäck, J. (2012). Election news in Sweden and the United States: A comparative study of sources and media frames. *Journalism*, 13 (5), 604–19. <https://doi.org/10.1177/1464884911431546>
- Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4 (6), 700–22. <http://doi.org/10.1080/21670811.2015.1096748>
- Doucette, C. (2016, December 22). Cops target 'pathetic parasites' who allegedly went on robbery spree. *Toronto Sun*. Retrieved from:
<http://torontosun.com/2016/12/22/cops-charge-pathetic-parasites-with-robberies/wcm/fo1cb6ao-e160-4c2d-9a3b-234coa9794f1>
- Ekbal, A., Saha, S., & Sikdar, U. K. (2013). Biomedical named entity extraction: Some issues of corpus compatibilities. *SpringerPlus*, 2 (ii), 601.
<https://doi.org/10.1186/2193-1801-2-601>

- Entman, R. (2010). Framing media power. In D'Angelo, P. & Kuypers, J. A. (Eds.), *Doing news framing analysis: Empirical, theoretical, and normative perspectives* (1st ed.) (pp. 331–356). New York, NY: Routledge.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43 (4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Entman, R. M., Matthes, J., & Pellicano, L. (2010). Nature, sources and effects of news framing. In Wahl-Jorgensen, K. & Hanitzsch, T. (Eds.). *The Handbook of Journalism Studies* (pp. 175–188). New York, NY: Routledge.
- Ethical journalism: A handbook of values and practices for the news and editorial departments. (2018). *New York Times*. Retrieved from: <https://www.nytimes.com/editorial-standards/ethical-journalism.html#introductionAndPurpose> on March 24, 2018.
- Fahmy, S., & Eakin, B. (2013). High drama on the high seas: Peace versus war journalism framing of an Israeli/Palestinian-related incident. *International Communication Gazette*, 76 (1), 86–105. <https://doi.org/10.1177/1748048513504046>
- Fox, C. (2016, December 22). Police say group of ‘pathetic parasites’ is responsible for 50 robberies since May. *CP24*. Retrieved from <http://www.cp24.com/news/police-say-group-of-pathetic-parasites-is-responsible-for-50-robberies-since-may-1.3213763>
- Franklin, B. & Carlson, M. (2010). *Journalists, sources, and credibility: New perspectives*. Abingdon, England: Taylor & Francis.
- Friend, C., Challenger, D., & McAdams, K. C. (2003). *Contemporary editing* (2nd ed.). New York, NY: Routledge.
- Gamson, W. A. (1992). *Talking politics*. Cambridge, MA: Cambridge University Press.
- Gans, H. J. (1979). Deciding what's news: A study of CBS Evening News, NBC Nightly News, Newsweek, and Time. In *Deciding what's news* (pp. 39–69). Evanston, IL: Northwestern University Press.
- Gillis, W., & Rankin, J. (2015, August 28). Critics see problems with carding review. *Toronto Star*. Retrieved from

<https://www.thestar.com/news/crime/2015/08/28/critics-see-problems-with-carding-review.html>

Gillis, W., & Rankin, J. (2015, December 5). Anti-carding coalition urges a bolder rewrite of Ontario's rules for 'street checks.' *Toronto Star*. Retrieved from <https://www.thestar.com/news/gta/2015/12/05/anti-carding-coalition-urges-a-bolder-rewrite-of-ontarios-rules-for-street-checks.html>

Gynnild, A. (2013). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, 15 (6), 72. <http://doi.org/10.1177/1464884913486393>

Ho, K. K. (2018). Marshall Project reveals diversity of its staff in new report. *Canadian Journalism Review*. Retrieved from <https://www.cjr.org/analysis/marshall-project-diversity-report.php>

Houston, J. B., Pfefferbaum, B., & Rosenholtz, C. E. (2012). Disaster news: Framing and frame changing in coverage of major U.S. natural disasters, 2000–2010. *Journalism & Mass Communication Quarterly*, 89 (4), 606–23. <https://doi.org/10.1177/1077699012456022>

Journalism code of ethics and professional practices. (2018). *Calgary Journal*, retrieved from <http://www.calgaryjournal.ca/pdf/EthicsCode09.pdf> on March 24, 2018.

Karlsson, M. (2011). The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority. *Journalism*, 12 (3), 279–95. <http://doi.org/10.1177/1464884910388223>

Karlsson, M., & Strömbäck, J. (2010). Freezing the flow of online news. *Journalism Studies*, 11 (1), 2–19. <http://doi.org/10.1080/14616700903119784>

Kellstedt, P. M. (2003). *The mass media and the dynamics of American racial attitudes*. New York, NY: Cambridge University Press.

Kennis, A. (2009). Synthesizing the indexing and propaganda models: An evaluation of U.S. news coverage of the uprising in Ecuador, January 2000. *Communication and Critical/Cultural Studies*, 6 (4), 386–401. <https://doi.org/10.1080/14791420903335127>

- Kovach, B., & Rosenstiel, T. (2014). *The elements of journalism: What newspeople should know and the public should expect* (3rd ed.). New York, NY: Three Rivers Press.
- Lawrence, R. G. (2010). Researching political news framing: Established group and horizons. In D'Angelo, P. & Kuypers, J. A. (Eds.) *Doing news framing analysis: Empirical, theoretical, and normative perspectives* (1st ed.) (pp. 265–86). New York, NY: Routledge.
- Lee, S. T., & Basnyat, I. (2013). From press release to news: Mapping the framing of the 2009 H1N1 A influenza pandemic. *Health Communication, 28* (2), 119–32.
<https://doi.org/10.1080/10410236.2012.658550>
- Lindgren, A. (2015). Aiding and abetting: How police media-information units shape local news coverage. In Richardson, C. & Smith Fullerton, R. (Eds.). *Covering Canadian crimes: What journalists should know and the public should question* (pp. 193–216). Toronto, Canada: University of Toronto Press.
- Livingston, S., & Bennett, W. L. (2003). Gatekeeping, indexing, and live-event news: Is technology altering the construction of news? *Political Communication, 20* (4), 363–80. <https://doi.org/10.1080/10584600390244121>
- Maat, H. P., & de Jong, C. (2012). How newspaper journalists reframe product press release information. *Journalism, 14* (3), 348–71.
<https://doi.org/10.1177/1464884912448914>
- Mahfouz, A. R. (2013). A critical discourse analysis of the police news story framing in two Egyptian newspapers before January 25 revolution. *European Scientific Journal, 9* (8), 309–33.
- Manning, P. (2001). *News and news sources: A critical introduction*. London, England: Sage Publications.
- Marrero, M., Sánchez Cuadrado, S., Morato Lara, J., & Andreadakis, G. (2009). Evaluation of named entity extraction systems. In Gelbukh, A. (Ed.) *Advances in Computational Linguistics, Research in Computing Science, 41*, 47–58.
- Matthes, J. (2009). What's in a frame? A content analysis of media framing studies in the world's leading communication journals, 1990–2005. *Journalism & Mass*

Communication Quarterly, 86 (2), 349–67.
<https://doi.org/10.1177/107769900908600206>

Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58 (2), 258–79.
<https://doi.org/10.1111/j.1460-2466.2008.00384.x>

McFarlane, D. J. (2011). Computational methods for analyzing health news coverage. *ProQuest LLC*. <https://doi.org/10.7916/D8571ZJ>

McGillivray, K. (2016, December 22). Group of ‘pathetic parasites’ arrested after string of gang-related robberies across GTA. *CBC News*. Retrieved from
<http://www.cbc.ca/news/canada/toronto/gang-related-robbery-1.3909023>

McNair, B. (2009). Journalism and democracy. In Wahl-Jorgensen, K. & Hanitzsch, T. (Eds.). *The handbook of journalism studies* (pp. 237–249). Portland, ME: Ringgold.

Miller, M. M. (1997). Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review*, 15 (4), 367–78.
<https://doi.org/10.1177/089443939701500403>

O’Neill, D., & O’Connor, C. (2008). The passive journalist. *Journalism Practice*, 2 (3), 487–500. <http://doi.org/10.1080/17512780802281248>

Otis, D. (2015, February 22). Police radio to go silent as Toronto cops move toward encrypted communications. *Toronto Star*. Retrieved from
<https://www.thestar.com/news/gta/2015/02/22/police-radio-goes-silent-as-toronto-cops-move-toward-encrypted-communications.html>

Owens, L. C. (2008). Network news: The role of race in source selection and story topic. *Howard Journal of Communications*, 19 (4),
<https://doi.org/10.1080/10646170802418269>

Perkins, K., & Kieffer, B. (2016, August 24). Journalistic values, objectivity, fairness, balance more important than ever. *Iowa Public Radio*. Retrieved from
<http://iowapublicradio.org/post/journalistic-values-objectivity-fairness-balance-more-important-ever#stream/o>

- Plaisance, P. L. (2007). Transparency : An assessment of the Kantian roots of a key element in media ethics practice. *Journal of Mass Media Ethics*, 22 (2-3), 37-41. <https://doi.org/10.1080/08900520701315855>
- Poindexter, P. M., Smith, L., & Heider, D. (2003). Race and ethnicity in local television news: Framing, story assignments, and source selections. *Journal of Broadcasting & Electronic Media*, 47 (4), 498-523. https://doi.org/10.1207/s15506878jobem4704_3
- Rankin, J., & Winsa, P. (2012, March 9). Known to police: Toronto police stop and document black and brown people far more than whites. *Toronto Star*. Retrieved from https://www.thestar.com/news/insight/2012/03/09/known_to_police_toronto_police_stop_and_document_black_and_brown_people_far_more_than_whites.html
- Rauhala, A., Albanese, P., Ferns, C., Law, D., Haniff, A., & Macdonald, L. (2011). Who says what: Election coverage and sourcing of child care in four Canadian dailies. *Journal of Child and Family Studies*. 21 (1), 95-105. <https://doi.org/10.1007/s10826-011-9481-0>
- Russial, J. (2004). *Strategic copy editing*. New York, NY: The Guildford Press.
- Scheufele, B. T., & Scheufele, D. A. (2010). Measuring frames and framing: Conceptual distinctions and their operational implications for framing research. In D'Angelo, P. & Lakatos, I. (Eds.). *Doing news framing analysis: Empirical, theoretical, and normative perspectives*. New York, NY: Routledge.
- Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society*, 3 (2-3), 297-316. https://doi.org/10.1207/S15327825MCS0323_07
- Schudson, M. (2003). *The sociology of news*. New York, NY: Norton.
- Shah, D. V., Cappella J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659 (1), 6-13. <https://doi.org/10.1177/0002716215572084>
- Sigal, L. V. (1973). *Reporters and officials: The organization and politics of newsmaking*. Lexington, MA: D. C. Heath.

- Sjøvaag, H., & Stavelin, E. (2012). Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence: The International Journal of Research into New Media Technologies*, 18 (2), 215–29. <https://doi.org/10.1177/1354856511429641>
- Soloski, J. (1989). News reporting and professionalism: Some constraints on the reporting of the news. *Media, Culture & Society*, 11 (2), 207–28. <https://doi.org/10.1177/016344389011002005>
- Spratt, M., Ferrand Bullock, C., Baldasty, G., Clark, F., Halavais, A., McCluskey, M., & Schrenk, S. (2007). News, race, and the status quo: The case of Emmett Louis Till. *Howard Journal of Communications* 18, 169–92. <https://doi.org/10.1080/10646170701309924>
- Strömbäck, J., Negrine, R., Hopmann, D. N., Jalali, C., Berganza, R., Seeber, G. U. H., ... Maier, M. (2013). Sourcing the news: Comparing source use and media framing of the 2009 European parliamentary elections. *Journal of Political Marketing*, 12 (1), 29–52. <https://doi.org/10.1080/15377857.2013.752227>
- Sylvester, E. (2016). Silenced no more. *Ryerson Review of Journalism*. Retrieved from <http://rrj.ca/silenced-no-more/>.
- Tankard, J. W. (2001). The empirical approach to the study of media framing. In Reese, S. D., Gandy Jr., O. H., & Grant, A. E. (Eds.). *Framing public life: Perspectives on media and our understanding of the social world* (p. 416). Mahwah, NJ: Lawrence Erlbaum Associates.
- Toronto Police Service. 2016. Corporate Communications. Retrieved from <http://www.torontopolice.on.ca/corpcomm/>.
- Van Gorp, B. (2009). Strategies to take subjectivity out of framing analysis. In D'Angelo, P. & Kuypers, J.A. (Eds.). *Doing news framing analysis: Empirical, theoretical, and normative perspectives* (1st ed.). (pp. 84–109). New York, NY: Routledge.
- Van Leuven, S., Deprez, A., & Raeymaeckers, K. (2014). Towards more balanced news access? A study on the impact of cost-cutting and web 2.0 on the mediated public sphere. *Journalism*, 15 (7), 850–67. <https://doi.org/10.1177/1464884913501837>

Ward, B. (2015, December 31). Need a newsroom resolution for 2016? How about reclaiming some expertise? *Poynter Institute*. Retrieved from <http://www.poynter.org/2015/need-a-newsroom-resolution-for-2016-how-about-reclaiming-some-expertise/390379/>

Zamith, R., & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What Computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, 659 (1), 307–18. <https://doi.org/10.1177/0002716215570576>