

Tracing Back Beyond the Text: Using Regressive Content Analysis to Study Journalistic Practices

Rastreando más allá del texto: el uso del análisis de contenido regresivo para estudiar las prácticas periodísticas

Dawn Wheatley. Dublin City University (Ireland)

Assistant Professor in the School of Communications focusing on journalism, political communication, and social media. PhD in 2018 for her research into online news production and coverage of the Irish healthcare sector, with a particular emphasis on sources and voices, and how newsroom practices are established and reproduced. Production journalist for the Irish Daily Mail and the Irish Times before turning to research and teaching in DCU. Her research has been published in *New Media & Society*, *Digital Journalism*, *Journalism Practice*, and on RTÉ's *Brainstorm*, and she teaches both practical and theoretical subjects to journalism students in the School. ORCID: <https://orcid.org/0000-0002-8751-4535>

Article received: 04/12/2020 – Accepted: 11/02/2021

Abstract:

Content analysis is a common tool in media studies and beyond, typically used to identify patterns in news texts to help inform study what and how information is communicated to the public. Yet such results, which are typically quantified based solely on the messages and details within the news report, have limitations for those researchers who seek to make a connection between journalistic output – the typical final news report – and the journalistic practices involved in the report's assembling. The methodology proposed here seeks to establish a systematic approach to exploring journalistic sourcing practices through the in-depth analysis of news reports, crucially by extending the research lens beyond the news report and exploring source contributions such as original press releases or other media reports. By systematically tracing back and recording such material, valuable and original findings can be made, which can then be cross-tabulated with results from the more traditional quantitative content analysis to provide fresh insights. This method requires two phases of research, outlined in this paper. The article also explores how content analysis can be used to look into journalistic practices, and a framework for conducting research, with advice and guidelines to both obtaining and recording the source material and data.

Keywords:

Content analysis, ethnographic, journalism, practices, qualitative, news routines, news sources

Resumen:

El análisis de contenido es una herramienta habitual en los estudios sobre los medios de comunicación y en otros ámbitos relacionados. Suele utilizarse para identificar patrones en los textos periodísticos que ayuden a estudiar qué y cómo se comunica la información al público. Sin embargo, estos resultados, que suelen cuantificarse basándose únicamente en los mensajes y detalles del informe de noticias, tienen limitaciones para los investigadores que buscan establecer una conexión entre la producción periodística -el típico informe final de noticias- y las prácticas periodísticas implicadas en el montaje del informe. La metodología que aquí se propone pretende establecer un enfoque sistemático para explorar las prácticas periodísticas de abastecimiento a través del análisis en profundidad de los reportajes, fundamentalmente ampliando el objetivo de la investigación más allá del reportaje y explorando las contribuciones de las fuentes, como los comunicados de prensa originales u otros informes de los medios de comunicación. Al rastrear y registrar sistemáticamente ese material, se pueden hacer descubrimientos valiosos y originales, que luego se pueden cruzar con los resultados del análisis de contenido cuantitativo más tradicional para proporcionar nuevas perspectivas. Este método requiere dos fases de investigación, que se describen en este artículo. El artículo también explora cómo puede utilizarse el análisis de contenido para estudiar las prácticas periodísticas, así como un marco para llevar a cabo la investigación, con consejos y directrices tanto para obtener como para registrar el material de origen y los datos.

Palabras clave:

Análisis de contenido, etnográfico, periodismo, prácticas, cualitativo, rutinas periodísticas, fuentes de noticias

1. Introduction

Content analysis has been a well-established method to explore a whole host of patterns and trends in media coverage, such as framing, bias, and gender portrayals, to name just a few. For some researchers within the journalism studies field in particular, content analysis has also provided an insight into the source material used in news reports, highlighting practices such as analysing the reliance on official sources (Brown et al., 1987) and official channels (Ciboh, 2017), and exploring what voices and actors are included, and excluded, in news reports (O'Neill & O'Connor, 2008; Ross, 2007). Methodologically, journalism researchers focusing on sources have used different approaches to gain insights into journalistic activity around the use of specific source material: for some, they start with the published content and trace backwards looking for public relations (PR) information (Lewis et al., 2008) while others start with press releases and analyse how they progress through newsrooms and news publications (Lee & Basnyat, 2013). Some have focused on the use of wire copy in news cycles (Johnston & Forde, 2011), or how social media content is incorporated into news reports (Broersma & Graham, 2013). When considering approaches which involve looking back to early reporting processes, the journalists themselves have also been a fruitful source of insight: Albaek (2011), in looking at how experts end up contributing to news content, revisited both the journalist-author and the expert to record the starting point/interactions. The reconstructive interview with journalists is also a route used to

establish what activity took place (Reich, 2006, 2014), relying on the journalists to account for their behaviours. However, despite these various methods, what is often lacking are more holistic approaches to thoroughly capture media content origins beyond individual categories of source material based solely on the media report, or turning to the journalists themselves, which may not always be feasible or appropriate.

Therefore, linking practice to published content requires a fresh approach to content analysis, one which moves beyond a one-dimensional analysis of the text which has limitations when trying to understand practice, discussed further below. For that reason, novel methods are needed to develop the link between daily actions and output: the approach outlined here is the basis of studies published elsewhere (Wheatley 2017, 2020a, 2020b), and this space provides an opportunity to focus more deeply on the steps and motivation of this approach. To strengthen the content analysis work into online journalism specifically, Steensen has suggested that it should be combined with more qualitative approaches to the analysis of texts which can help to understand the field's complexity (2011, p. 322). The method proposed here is in no way limited to online journalism, even though that is the source material which inspired its development, but may work best for text-based research when passages of text can be easily extracted and searched for elsewhere.

Previous work on qualitative content analysis has thoroughly explored its merits (Mayring, 2014), but not always with a clear focus on methods appropriate to studying journalism content and, therefore, this paper contributes to the field by providing such an approach. This article begins with a discussion about content analysis and how it can be used to analyse and deconstruct sourcing practices, despite some researchers' reservations. The focus then shifts to the content analysis method proposed here. Firstly, the two-phase content analysis, which was needed to capture elements of both the sourcing and publication process, is described, which firstly involves recording the manifest data of the sample and the use of sources in each story, and the second draws on the more qualitative document analysis which Altheide (1987) describes. The overview of this method of analysis, branded here as "regressive content analysis", includes detail about its key steps, and how its potential was enhanced by using code-and-retrieve qualitative analysis software (NVivo) to maximise analytical outcomes.

2. Using content analysis to investigate journalistic practices

Content analysis as a research method is based on the assumption that the material being analysed is a fair reflection of the behaviour and values of those who create it (Berger, 1998, p. 23), while Kolmer even argues that in terms of news production, it is not possible to assess journalists' daily work without reference to the output and final product (2008, p.117). They point out that the qualities of published content enable researchers to consider the relevance of the cultural, political and economic framework for the production of media content. Meanwhile, Weaver (2008, p. 107) believes observations and content analyses are more valid and reliable methods of studying what journalists do rather than survey data, while Berger (1998, p. 26) points out that the unobtrusive nature of content analysis increases its reliability, which may be a concern with observational studies (Neuman, 2014).

One advantage of content analysis studies is that, according to Schröder, “they can serve to confirm or disconfirm intuitive impressions” through systematic descriptions of large sets of media discourses (2002, p. 102). He adds that a “trade-off” is the “inevitable reduction of complexity that follows from the decontextualisation of meaningful elements”, a concern Hansen (1991) also raises. However, such a trade-off may more often be associated with the purely quantitative, statistical content analysis approaches in which advocates seek a clear detachment (Krippendorff, 2013). Focusing on content analysis and sources specifically, Hansen et al (1998) refer to one criticism of quantitative content analysis: that frequency of occurrence does not explain influence. Drawing on the terms of primary definition of sources in news, echoing Hall et al (1978), they say that, to understand such sources’ weight, “it is necessary to take into consideration the ‘newsmaking scenarios’ or ‘fora’ through which such primary definers become newsworthy and articulate their claims (ibid., p.110).

These ‘newsmaking scenarios’ are central to many studies into journalistic practices but are often underplayed or not fully considered in traditional content analysis. However, establishing a clear link between process and output raises the argument made by some researchers that content analysis alone cannot be used to explain practice. Boesman et al (2015) and Reich (2006) advise that content analysis alone is insufficient for tracing the origins of news content. Boesman et al suggest: “The appearance of routine sources in a news story does not automatically mean that the story was found through a routine newsgathering channel”, and indicate that content analysis alone is inappropriate for production studies (2015, p. 909). This echoes Reich (2006, p. 501), who argues that content analysis is insufficient as it examines the product, not the process. He suggests that any effort to link production to process rests on two “unfounded assumptions”: that news processes are encoded into the content, and that researchers can decode them reliably. However, it is this limitation that the two-phase, regressive content analysis proposed here seeks to overcome. Both Boesman et al and Reich cite Manning (2001, p. 48), who highlights the limitations of focusing solely on media texts. However, when Manning’s original work is examined, there may be scope for more thorough content analysis – moving beyond the media text – to be used to consider processes and power. Manning writes:

Unless we begin to consider the news encoding process and news source strategies associated with the production of a news text, and the power relations underpinning these, the “hierarchies of definitional power” are left in a vacuum, located in the text but lacking a social context. (ibid.)

Manning refers to the “difficulties inherent in trying to make inferences about social practices or power relations from the evidence of media texts alone”, rather than actually studying the encoding process itself. He adds:

There has certainly been an explosion in the number and intensity of the information flows circulating in society but there is also some control and some order in the way they enter both the public and private domains. The task of a sociology of news and news sources is to trace the sources of order and control without reducing or essentialising a complex social reality. (ibid.)

The reference to news source strategies, practice and the “order” in how information enters the public domain can be clearly linked to the channels through which news originates. Yet to explore such channels, information from outside of the media text must be examined. Therefore, attempting to capture some of this complexity can be done through content analysis but may require additional approaches which expand beyond the news text. One of the limitations Hansen put on her study of sources three decades ago was that “the content analyst is not able to detect information subsidies provided to the media organisations” (1991, p. 482), which resonates with Reich’s concern about the researcher’s inability to decode reliably. However, that is no longer the barrier it once was: online search tools which are pervasive in the contemporary digital landscape facilitate information-sourcing and make it easier than ever to seek and obtain source material. Bertrand and Hughes have warned against making unwarranted inferences based on content analysis data alone, and “taking the discussion further than the data legitimately allows” by basing inferences solely on content analysis (2005, p. 179). Yet, in the novel regressive approach proposed here, the fact that the analysis process includes material which is drawn from beyond the news text actually strengthens any inferences made about how source material and source actors interact, thus forming a link between practice and output, or process and product.

For many researchers content analysis appears as a purely quantitative activity to gather statistical data, measuring the extent and frequency of messages (Sumser 2000, p.209), but there is the potential for content analysis to become much more qualitatively focused. Priest (2010) makes the distinction between “relatively superficial” manifest content, and “deeper, symbolic” latent content: qualitative content studies can “better take into account subtleties of the structure of arguments and narratives not easily captured by quantitative summaries” (ibid., p.108). While most news content analysis studies are, by their nature, restricted to the text or broadcast, and generate quantifiable results, the actual coding process in Phase 2 of this method extends beyond the text. Rather than considering this content analysis as either quantitative or qualitative, the approach outlined here draws on Krippendorff’s questioning of the validity of such distinctions: “Ultimately all reading of texts is qualitative, even when certain characteristics of a text are later converted into numbers” (2013, p. 22). This is especially apparent as many of the results garnered through the regressive stage can be ultimately quantified.

Krippendorff (2013) lists a number of alternative protocols that are more qualitative in nature, which includes the approach advocated by Altheide known as *ethnographic content analysis*, which seeks a more iterative, reflexive approach as the analysis process develops. Krippendorff, however, warns that “unfortunately ... this call has not yet yielded specific methods of analysis” (2013, p. 27). Such a limitation and lack of specific protocols and precedents on which to draw meant that new tools were needed for the approach proposed here; the following section presents a brief overview of Altheide’s approach and outlines how it was applied here.

3. Source material and an alternative approach to analysing content

Altheide (1987; 1996) has written extensively about the potential for gathering large amounts of qualitative material through what he calls *ethnographic content analysis* (ECA), or qualitative document analysis—Altheide seemingly refers to “ethnographic

content analysis” (ECA) and “qualitative document analysis” interchangeably in his book *Qualitative Media Analysis* (1996)—. He says the context or social situations surrounding a document must be understood, particularly in the case of news reports:

They are organisational products. This again suggests the importance of process, or how something is actually created and put together. Newspapers and TV newscasts are put together according to a routine and a complex division of labour and deadlines. (Altheide 1996, p.9)

The nature of how the information used during the coding process is obtained and recorded is a move away from more traditional content analyses. While Altheide refers to the qualitative term, it is not a textual or discourse analysis; instead, it captures the distinction between quantitative content analyses and document analysis, which requires a more inductive, reflexive method. Altheide and Schneider outline the key differences between ECA and quantitative content analysis (QCA), reproduced in Table 1, such as reflexive research, purposive sampling and the collection of narrative data. Similar steps and stages are also proposed elsewhere (Mayring, 2014), again emphasising the preparatory steps and the need for each approach to be tailored to each specific study.

Table 1

The differences between quantitative content analysis and ethnographic content analysis

	Quantitative content analysis	Ethnographic content analysis
Research goal	Verification	Discovery; Verification
Reflexive research design	Seldom	Always
Emphasis	Reliability	Validity
Progression from data collection; analysis, interpretation	Serial	Reflective; circular
Primary researcher involvement	Data analysis and interpretation	All phases
Sample	Random or stratified	Purposive and theoretical
Pre-structured categories	All	Some
Training required to collect data	Little	Substantial
Type of data	Numbers	Numbers; Narrative
Data entry points	Once	Multiple
Narrative description and comments	Seldom	Always
Concepts emerge during research	Seldom	Always
Data analysis	Statistical	Textual; Statistical
Data presentation	Tables	Tables and text

Note. Reproduced from Altheide & Schneider (2013, p.3, 24).

Therefore, taking an ECA approach to documents offers the opportunity to collect quantitative data, as well as narrative descriptions and textual analysis. Altheide says this flexibility, interactive and reflexive approach contrasts with the positivist contentions of quantitative content analysis. He presents 12 stages of the gathering and coding process for using ECA and the stages broadly mirror the analytical progression of this study which inspired this method, and capture the reflexive approach to the process, illustrated in Table 2.

Table 2

The process of qualitative document analysis, including how each stage was used in this original study

Stage (Altheide 1996)	Application to this study
1. Topic	Select the subject area
2. Ethnographic study/lit	Observe the news product to establish ideas
3. A few documents	Pilot period
4. Draft protocol	Pilot coding sheet
5. Examine documents	Pilot coding process
6. Revise protocol	Revision of coding
7. Theoretical sample	Decide on specifics of sample
8. Collect data	Data gathered over 14 weeks
9. Code data	Stories coded with coding sheets and in NVivo software
10. Compare items	Run queries and analyse data
11. Case studies	Selected based on suitability factors garnered during coding
12. Report	Results and analysis discussion

Note. Adapted from Altheide 1996, p.13

Crucial to this approach is how each stage informs the next: Altheide says the aim is to be “systematic and analytic, but not rigid” (ibid., p.16). The need to be systematic in qualitative work is also highlighted by Sumser, who advises anybody doing any kind of content analysis to think in variables as much as possible to ensure the research is empirically sound, “as the burden of proof needed is often higher in qualitative research” (2000, p. 209). A similar point is put forward by Mayring, who argues that attempts should be made to “concretise the objectives of the analysis in category form”, thus becoming the central instrument of analysis (2014, p. 40). Of course, variables will depend on the study and data in question, and the project’s aims, objectives and research questions.

Elsewhere, one point to consider in selecting methodologies is the suitability for the researcher: Kolmer (2008, p.120) says a certain level of circumstantial knowledge is always necessary to understand a text correctly. For Altheide, “ideally” the researcher

using ethnographic content analysis will have a general awareness of news production methods and be familiar with the “context and process... to adequately consider the relevant aspects of a news report, including the knowledge to rule out erroneous explanations of news content” (1996, p.9). Therefore, any study of news production or sourcing may benefit from a lead researcher who is familiar and confident in understanding some of the mechanics of how stories are assembled and how sources behave and interact with journalists. Regarding Table 2, many of the early steps are ultimately preparatory and involve establishing, through reflexive thinking and an open mindset, the parameters of the study and its analytical scope. For the purpose of this article, though, most attention is given to steps 8-10, where the specific tools and processes are most involved, further detailed below.

4. Methodology: two-phase coding process

Phase 1: Traditional content analysis with codebook

Traditional quantitative content analysis will typically have a deductive codebook in place, whereby the questions and coding protocols are clearly categorised and answers can be easily quantified. The codes and questions vary depending on the scope of the study, but may include details such as publication, byline, dominant news value, frames, gender aspects, or many other elements. In the original study which inspired this methodological approach, started in this way, with the first phase recording attributes about the story’s publication (publication, day of the week, time of publication, authorship type, sex of author, number of sources used, the first source used, sex of sources used). These first-phase results are generated by looking solely at what appears in the text, and following the guidelines and practices advised in content analysis best practice, tailored to the particulars of the projects and its objectives.

Phase 2: Regressive content analysis

The second phase is where the coding shifts beyond the text itself, and instead the researcher seeks out some of the contributing material and sources used in the story’s assembling. The key question here was for every source contribution observed: where did the contribution come from? (This question may, of course, vary depending on the particular study’s goals). Answering this involved tracing back the source material, where possible, and recording the passages of text in the story and where each originated.

For this second phase of analysis, the stories were imported into the code-and-retrieve NVivo software (further outlined in the following section) and each story was treated as an individual unit, known as a “node” in the software. Drawing on a dual categorisation, each story was coded – via the NVivo method of highlighting and recording relevant text – for two elements: (A) The sources and perspectives used in each story and (B) the origins of the content. For example, quotes from a trade union leader that were traced to a press release were recorded both as (A) “trade union”, and (B) “press release”. Similarly, quotations from a parliamentary speech by a politician were coded as both (A) “politician” and (B) “speaking in parliament”. Ultimately, each source contribution included one code from category A, and one code from category B, as outlined in Table

3. Depending on the study’s topic and purpose, the list of contributors and source groups will vary and should be informed during the pilot process.

The items in this original study totalled 896 news reports from five mainstream and popular daily news websites in the Republic of Ireland, gathered over 14 weeks. The stories were focused around daily coverage of the health service and health policy, so included topics such as the running of hospitals, waiting lists, access to drugs, hospital and treatment waiting lists, labour disputes among doctors and nurses, NGOs lobbying for funding for specific conditions, and human interest stories documenting individuals’ experiences within the Irish health system. As the stories under analysis related to healthcare policy, the key source categories included politicians, private citizens, state agencies, trade unions, medical professionals, hospital management, and business owners.

Table 3

The dual categorisation for coding source material in news reports

A: Source (voice/actor) Code the source contribution to one of the following:	B: Source (origin of content) Code the source contribution to one of the following:
Politician	Clearly original material
State agency	Information subsidy
Union/lobby group	PR statement in response to issue
Private citizen	Report
Medical professional	Social media content
Hospital management	Other media (eg radio, newspaper)
Civil society/voluntary group	Leaked information
NGO/charity	Freedom of Information
Commercial organisation/ businessperson	Media interview/press conference
Anonymous	Live planned event
Other	Academic research
	Parliament/parliamentary committee
	Unknown origins
	Other

Category A: Sources and perspectives

The first aspect was relatively straightforward process of assigning each text attributable to a source to a category. This included whenever the source was speaking, directly or indirectly, or if the information used was directly attributed to a source or gave their perspective. Only information clearly attributable to sources or actors was coded for this part; this meant not all text in a story was coded, as much of it was general descriptive information or background which could not be linked directly to a source or voice.

Category B: Origin of content and obtaining source material

While the source voices and perspectives were relatively straightforward to code, the origins of source material was more challenging.

Firstly, parameters were established to determine what elements of the news text would need to be traced back and this aligned with the previous parameters regarding source/voice contribution: any instance in which a statement, observation or detail was inferred or attributed was considered some sort of source material. This meant that not everything in the news text was coded, but typically approximately 60-70% of most news texts were built around these source contributions. During a pilot period this became one of the most important aspects to resolve and refine.

Once these parameters were established, the process of tracing back each segment began. To do this, various steps were taken to try to seek out the original source material, with the most common routes outlined below.

In-text attribution: The original source was sometimes mentioned, such as “in a statement”, or “speaking to the” publication in question. This was particularly helpful for what were considered “original/exclusive” contributions obtained by the journalist (which were coded as such), which are, understandably, one of the most difficult exchanges to accurately record.

Online search engines: This was the most used channel through which original material was obtained. Taking a quotation from the news report and searching for it via a search engine was the most effective way (as quotations were unlikely to be reworded by the journalist), or any other distinctive detail that might lead to search results. Although not recorded systematically during the process, an estimated 50% of source contributions were obtained this way, with most results directing to information subsidies of various forms on stakeholders’ websites.

Stakeholder websites: While still not fully transparent, the development of organisations’ websites and media strategies mean that much of the information provided to journalists is now easily attainable for a researcher. For example, reports commissioned by NGOs or press releases from trade union websites were often easily available, and it was possible to account for much of the source material this way. Some of this was obtainable through search engines, and while this provided text-matches to the content, finding the full original information subsidies – and how the news angles were challenged or replicated by journalists – also provided some key narrative detail to record, which is so important in qualitative content analysis.

Social networks: If there was a clear contribution from a named individual in the news story, but search engines and in-text references were not making its origins apparent, it was also useful to search for the individual on social networks, and Twitter in particular. The running-commentary nature of a social network like this means that users will often share their thoughts about individuals who might be appearing at that moment: on a radio show, for example, or making a speech. Furthermore, given many journalists’ heavy use of Twitter, it also provides insights into the journalists’ daily activities and where they might encounter various sources and events which can also act as a prompt.

Elsewhere, this was particularly effective for smaller, regional radio stations whose results might not appear on websites, but their promotional Twitter account might refer to the fact a certain guest was appearing on a show. Recordings of the radio show could then be sought to confirm the quoted contributions.

Other media: The prevalence of news organisations repackaging or “cannibalising” content from other media outlets was one of the starkest results from the content analysis, an understanding of which was only made possible because of this methodology. As with the above categories, sometimes the news text included references to the original publication/broadcast, or details might show up in search results, but in other instances, the original content was sought. Although specific to the country under investigation, the website transcripts.ie (no longer operational) was useful in providing a comprehensive (but not fully reliable) log of guests who had appeared on major national radio programmes. A similar service, TVEyes (available through Lexis Nexis database), offers details on TV and radio guests on the main programmes in the UK and other such services may be present available elsewhere.

Parliament: Searching the parliamentary website was also helpful for transcripts from the two main chambers and various parliamentary committees. These results would often not show up on search engines but provided a host of source content used by journalists.

Unknown: Of course, not all material could be traced and unobtainable content should be recorded as such. In the original study described here, if the source material could not be obtained, it was coded as “unknown origins”, which comprised 7.5% of the content.

5. NVivo software

The development and employment of this approach would not have been possible without the appropriate software to manage the study’s aims and the peculiarities of the methodology. After the Phase 1 database was complete (via a coding sheet created using Google Forms, and then exported to Excel), the data was imported into NVivo. NVivo is one of the most popular types of CAQDAS (Computer-Assisted Qualitative Data Analysis Software), designed to assist with a variety of data types, such as text, transcripts, photographs, audio and video. It combines a code-and-retrieve system and advanced search functions; Hoover and Koerber (2011) explain this further saying: “Code and retrieve allows the researcher to manually code snippets of the data according to their common themes. That coding can then be retrieved and viewed separately from the original data”, adding that it becomes particularly useful when one can then process and query the coding (2011, p. 71). Therefore, this is helpful for more inductive approaches to coding and recording data which requires more than a deductive coding form. The software adds efficiency to processing and analysing the data, and a level of transparency which is sometimes seen to be lacking in qualitative work. Hoover & Koerber (2011) say NVivo enhances their work in three key ways: efficiency, multiplicity and transparency. However, they note how it “does not do the intellectual work” and is a secondary tool in the research process:

Although the automated features allow some types of coding to be accomplished much more quickly, these features do not by any means replace the need for careful reading of every line of data ... Just as a strong statistical program like SPSS cannot make up for poorly gathered numbers, strong CAQDAS cannot make up for poorly gathered material or shallowly considered analysis. (Hoover & Koerber, 2011. pp.71/77)

Similarly, Zamawe (2015) makes the point that the main function of such software is not to analyse the data, per se, but to aid and support the analysis process by acting as a data management package, also echoed by Mayring (2014) speaking about qualitative software generally. However, the range of what NVivo (or other similar software) can help analyse, and the relationships between codes, is undoubtedly one of its strengths. Bazeley (2009) argues that qualitative analysis can go beyond simply identifying themes and describing patterns in discourse. She says that matrix displays can be displayed which are generated based on co-occurrence of codes, or of codes and the attributes associated with each node (in this study each story unit is a node, and an attribute was, for example, the publication or author type). Bazeley says this allows a researcher to “assess both patterns of association (how often things vary under different circumstances), and the nature of the associations (in what ways something might vary under particular or different circumstances)” (ibid., p.14). This becomes crucial when trying to link practices and output, as the various codes offer insight into not just the different voices being used, but the sourcing channels through which they are accessed by journalists.

6. Results & analysis of data

Using NVivo allowed the two phases of coding to be combined, and links found between, for example, the time of day a story is published, the dominant voices heard, and the origins of the material used in the news report. Such analysis can be carried out with any of the story attributes which were recorded (such as authorship type, specific publication), and combined with the origins of the material, or the use of particular voices. It, therefore, becomes possible to analyse each news text from various perspectives: publication details, journalistic sourcing practices, and the outcome, in the context of what voices are heard or omitted. When both coding stages were complete, each story node has extensive details of information linked to it, which formed the basis for the analysis. This included the manifest data, such as a publication and authorship details, and the number of sources, as well as the latent information from the second phase of coding voices present and origins of content. The analysis drew on both these data types and used pivot tables in the Microsoft Excel database and NVivo matrix queries. In NVivo, any query could be run with more than 40 variables available here to use alongside the qualitative coding of source material and voices used. Furthermore, the results in NVivo queries could be measured in either word count, the number of occurrences, or the number of stories, depending on the desired analysis and suitability.

Although in-depth analysis of the original study's results is beyond the scope of this article, it is perhaps appropriate to demonstrate some of the particulars of how the methodology provided valuable empirical insights into journalistic sourcing practices.

The study demonstrated the influence of health-related state agencies, politicians, and trade unions in gaining media attention via PR material which was typically repurposed without much amendments from the journalists. Elsewhere, we learnt that ordinary/private citizens were often included in the news and becoming salient voices, but often only because of their contributions to talk radio shows in which users call in and make comments, which then get repurposed by journalists in other rival outlets (Wheatley, 2020b). Regarding temporality and the rhythms of online publishing cycles, this methodological approach also allowed us to see that among legacy newspaper brand outlets, there was still often a print-first mentality still present; more original journalism work was often held for the print edition first, and more superficial reporting (often based on press releases and repackaged material from other media) was typically used to populate the outlets' websites during the day, helping to fill the perceived open-ended news gap (Wheatley & O'Sullivan, 2017). More generally, the approach also facilitated a conceptual framework building on previous work into channels of production (Sigal, 1973), to help researchers categorise news reports into two broad categories (routine and non-routine), containing eight subcategories: A) promotional PR; B) non-promotional PR; C) other media reports; D) predictable events; E) unexpected events; F) leaks; G) enterprise reporting; H) special reports/analysis (Wheatley, 2020a). These categories were refined and reconceptualised from the original "Origin of content" categories identified earlier in this article.

7. Conclusion

The mixed-method approach outlined here facilitates a novel means to explore how and from where stories are sourced and who gets heard, or excluded. The non-traditional approach to content analysis is required to establish the often-overlooked link between practice and published content, but when this original approach is used, invaluable insights can be garnered regarding journalist sourcing activities far beyond what traditional content analyses—limited to the text—can achieve.

The first phase of content analysis establishes the publication-related data, such as website, time, and authorship details; these attributes provides a crucial starting point from which online publishing activity can be investigated. However, much more data and deeper insight is needed to more fully understand journalistic sourcing practices. The second phase of coding, therefore, moves the investigation on to analyse the source material and record the voices used: this dual approach is intrinsic to any analysis of source use and to fully understanding how different groups are affected and gain a presence within the news environment.

Ultimately, this approach is made possible because of the availability online of much of the original source material upon which journalists draw. Of course, it must be acknowledged that this approach may not be fruitful in all such studies, with the availability of source material likely varying by beat: the stakeholders involved in certain topics may use online dissemination platforms more than those involved in certain other fields. This original study was based on news coverage of the health policy sector in which many of those involved are open in their stances, such as politicians, labour unions and patient representative groups, but this may not be replicable in fields in which there is more covert communication strategies at play. Nevertheless, there are many journalistic fields and studies in which the approach outlined here can be useful

and valuable to help researchers better understand an often-overlooked dimension of the news sourcing practice.

8. References

- Altheide, D. L. (1987). Ethnographic Content Analysis. *Qualitative Sociology*, 10(1), 65–77.
- Altheide, D. L. (1996). *Qualitative Media Analysis*. Sage Publications.
- Altheide, D. L., & Schneider, C. J. (2013). Ethnographic Content Analysis. In *Qualitative Media Analysis* (pp. 23–37).
- Bazeley, P. (2009). Analysing Qualitative Data: More Than ‘Identifying Themes.’ *The Malaysian Journal of Qualitative Research*, 2(2), 6–22.
- Berger, A. A. (1998). *Media Research Techniques* (2nd ed). Sage Publications.
- Bertrand, I., & Hughes, P. (2005). *Media Research Methods: Audiences, Institutions, Texts*. Palgrave Macmillan.
- Boesman, J., D’Haenens, L., & Van Gorp, B. (2015). Triggering the News Story. *Journalism Studies*, 16(6), 904–922.
<https://doi.org/10.1080/1461670X.2014.953783>
- Broersma, M., & Graham, T. (2013). Twitter As a News Source. *Journalism Practice*, 7(4), 446–464. <https://doi.org/10.1080/17512786.2013.802481>
- Brown, J. D., Bybee, C. R., Wearden, S. T., & Straughan, D. M. (1987). Invisible Power: Newspaper News Sources and the Limits of Diversity. *Journalism & Mass Communication Quarterly*, 64, 45–54.
<https://doi.org/10.1177/107769908706400106>
- Ciboh, R. (2017). Journalists and Political Sources in Nigeria. *The International Journal of Press/Politics*, 22(2), 185–201.
<https://doi.org/10.1177/1940161216681164>
- Hall, S., Critcher, C., Jefferson, T., Clarke, J., & Roberts, B. (1978). *Policing the Crisis: mugging, the state, and law and order*. Macmillan.
<https://doi.org/10.2307/589668>
- Hansen, A., Cottle, S., Negrine, R., & Newbold, C. (1998). *Mass Communication Research Methods*. Macmillan.
- Hansen, K. A. (1991). Source Diversity and Newspaper Enterprise Journalism. *Journalism & Mass Communication Quarterly*, 68(3), 474–482.
<https://doi.org/10.1177/107769909106800318>
- Hoover, R. S., & Koerber, A. L. (2011). Using NVivo to Answer the Challenges of

- Qualitative Research in Professional Communication: Benefits and Best Practices Tutorial. *IEEE Transactions on Professional Communication*, 54(1), 68–82. <https://doi.org/10.1109/TPC.2009.2036896>
- Johnston, J., & Forde, S. (2011). The silent partner: News agencies and 21st century news. *International Journal of Communication*, 5(1), 195–214.
- Kolmer, C. (2008). Methods of Journalism Research - Content Analysis. In M. Loffelholz & D. Weaver (Eds.), *Global Journalism Research: Theories, Methods, Findings, Future* (pp. 117–130). Blackwell Publishing.
- Krippendorff, K. (2013). *Content Analysis: An Introduction to its Methodology* (Third). SAGE Publications.
- 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), 119–132. <https://doi.org/10.1080/10410236.2012.658550>
- Lewis, J., Williams, A., & Franklin, B. (2008). A Compromised Fourth Estate? UK news journalism, public relations and news sources. *Journalism Studies*, 9(1), 1–20. <https://doi.org/10.1080/14616700701767974>
- Manning, P. (2001). *News and News Sources: A Critical Introduction*. SAGE.
- Mayring, P. (2014). *Qualitative Content Analysis Theoretical Foundation, Basic Procedures and Software Solution*.
- Neuman, W. L. (2014). Social Research Methods: Qualitative and Quantitative Approaches. In *Social Research Methods: Qualitative and Quantitative Approaches* (Seventh). Pearson. <https://doi.org/10.2307/3211488>
- O'Neill, D., & O'Connor, C. (2008). The Passive Journalist. *Journalism Practice*, 2(3), 487–500. <https://doi.org/10.1080/17512780802281248>
- Priest, S. H. (2010). *Doing Media Research: An Introduction* (Second). Sage.
- Reich, Z. (2006). The Process Model of News Initiative: Sources lead first, reporters thereafter. *Journalism Studies*, 7(July), 497–514. <https://doi.org/10.1080/14616700600757928>
- Reich, Z. (2014). Islands of divergence in a stream of convergence: Comparing the news practices of male and female journalists in Israel. *Journalism Studies*, 15(1), 64–81. <https://doi.org/10.1080/1461670X.2013.790619>
- Ross, K. (2007). The journalist, the housewife, the citizen and the press: Women and men as sources in local news narratives. *Journalism*, 8(4), 449–473. <https://doi.org/10.1177/1464884907078659>
- Schrøder, K. C. (2002). Discourses of fact. In K. B. Jensen (Ed.), *A Handbook of Media and Communication Research* (pp. 89–116). Routledge.

- Sigal, L. V. (1973). Reporters and officials: the organization and politics of newsmaking. D. C. Heath.
- Steensen, S. (2011). Online Journalism and the Promises of New Technology. *Journalism Studies*, 12(3), 311–327. <https://doi.org/10.1080/1461670X.2010.501151>
- Sumser, J. (2000). *A guide to empirical research in communication: rules for looking*. Sage Publications Inc.
- Weaver, D. (2008). Methods of Journalism Research - Survey. In M. Loffelholz & D. Weaver (Eds.), *Global Journalism Research: Theories, Methods, Findings, Future* (pp. 106–116). Blackwell Publishing.
- Wheatley, D. (2020a). A Typology of News Sourcing: Routine and Non-Routine Channels of Production. *Journalism Practice*, 14(3), 277–298. <https://doi.org/10.1080/17512786.2019.1617042>
- Wheatley, D. (2020b). Victims and Voices: Journalistic Sourcing Practices and the Use of Private Citizens in Online Healthcare-system News. *Journalism Studies*, 21(8), 1017–1036. <https://doi.org/10.1080/1461670X.2020.1727355>
- Wheatley, D., & O’Sullivan, J. (2017). Pressure to Publish or Saving for Print? Digital Journalism, 5(8), 965–985. <https://doi.org/10.1080/21670811.2017.1345644>
- Zamawe, F. C. (2015). The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections. *Malawi Medical Journal*, 27(1), 13–15.

Conflict of Interest: The author declares that they have no conflict of interest

HOW TO CITE (APA 7^o)

Wheatley, D. (2021). Tracing Back Beyond the Text: Using Regressive Content Analysis to Study Journalistic Practices. *Comunicación y Métodos | Communication & Methods*, 3(1), 40-55. <http://doi.org/10.35951/v3i1.93>