Content Analysis

KIMBERLY A. NEUENDORF and ANUP KUMAR
Cleveland State University, USA

Introduction

Content analysis is perhaps the most widely used method in the academic field of political communication, increasingly in combination with other methods such as surveys, focus groups, and experiments. Since the 1950s, when Bernard Berelson introduced the empirical approach to the systematic study of "manifest content" in all forms of communication, its use has included the examination of political content in news media, speeches, advertisements, and campaigns, and recently social media and blogs. Content analysis includes text analysis, the systematic study of written text or transcribed speech, as well as techniques that focus on nontextual message content, including pictorial images, graphical elements, moving images, nonverbal behaviors, music, and sounds.

Content analysis has been defined as the systematic, objective, quantitative analysis of message characteristics. However, the utility of qualitative approaches is indisputable. The distinction between quantitative and qualitative analyses of message content is sometimes contested, varying between (a) whether the constructs of interest are principally quantitative or qualitative in nature, and (b) whether the measures of these constructs result in quantifications or more qualitative (either microscopic or holistic) descriptions of the messages.

Methodological issues in quantitative content analysis

Quantitative content analysis tends to follow the precepts of the scientific method, including attention to the standards of objectivity/intersubjectivity, reliability, validity, and generalizability (for comprehensive methods coverage, see Further reading). Two main methodological choices exist, coding by human coders (judges/raters), and computer coding (i.e., computer-aided text analysis, or CATA; e.g., Gottschalk & Bechtel, 2008). The two approaches are sometimes used in concert, but are also employed independent of one another.

Human coding

Most content analyses include some manual or human coding. General methods texts on content analysis provide guidance on the essentials, including theoretic backing, the construction of a coding scheme, coder training, and intercoder reliability assessment.

Other important issues include considerations of unitizing of content (i.e., how content is divided into identifiable units for coding), and substance versus form variables (substance characteristics are those that may appear or exist in any medium, while form features are relevant to the medium through which the message is sent). Additionally, consideration is given to the issue of whether content is manifest or latent; manifest content may be defined as elements that are present and directly identifiable, while latent content constitutes the deeper meaning, that not directly observable. Increasingly, scholars view the manifest/latent discrimination on a continuum, rather than as two discrete types.

A content analysis coding scheme for human coding is a written collection of measures with full explication, allowing for the careful application of the scheme by coders working independently, after sufficient training. The development of a coding scheme is a painstaking process, including pilot testing before finalization.

Computer-aided text analysis (CATA)

Beginning in 1965 with Philip Stone's General Inquirer, all CATA programs have as their basis the analysis of text via the application of algorithms of word/sequence searching and counting. Most often, the analysis involves one or more dictionaries (i.e., lists of search terms). CATA programs range from those that provide only word-count output, to those that include multiple provided dictionaries, to those that allow no dictionaries but instead focus on patterns of co-occurrences of words (see http://academic.csuohio.edu/neuendorf_ka/content; http://www.content-analysis.de/software; http://www.textanalysis.info/).

CATA programs typically include pre-set dictionaries derived from particular epistemological and theoretic underpinnings. Most salient for the political realm are LIWC (Linguistic Inquiry and Word Count), developed from a psycholinguistic perspective by James Pennebaker, and Diction, developed from a political-rhetorical tradition by Rod Hart. LIWC (http://liwc.net) features 84 dictionaries that tap such linguistic and semantic concepts as use of first-person pronouns, anger, optimism, and reference to motion. Diction (http://www.dictionsoftware.com), designed to analyze political speech, has 31 provided dictionaries, including those intended to measure tenacity, aggression, satisfaction, and complexity, which are also combined to form "master variable" scales: activity, optimism, certainty, realism, and commonality. The creation of custom dictionaries, a demanding process both conceptually and logistically, is allowed in most CATA programs, including the flexible freeware Yoshikoder (http://yoshikoder.org).

CATA has begun to be used for the assessment of "sentiment" or "tone" of political communication, including social media feeds expressing public opinion, and other online political content. For example, Pennebaker and Chung (2008) used LIWC to analyze the linguistic style of al-Qaeda leaders, identifying a post-9/11 increase in cognitive complexity, emotionality, and use of anger and hostility words in statements by Osama bin Laden. However, customized computer algorithms are the norm in this nascent but expanding adjunct to traditional content analysis (Liu, 2012), generally limiting the scope to positive/negative comments, and raising the issue of the validation of the classification of sentiment associated with words and phrases.

Reliability in content analysis

The reliability of a measure is the extent to which it produces the same result on repeated trials; in content analysis, this means the extent to which a measure in a coding scheme produces the same result when applied by different human coders. The assessment of intercoder reliability is crucial to human-coded content analysis.

Standard texts on the method of content analysis will address the conceptual issues of reliability, and provide practical advice on how to execute intercoder reliability checks, including the selection of a subsample for reliability assessment. One area of reliability in which researchers are forging new knowledge is with regard to changing options for intercoder reliability coefficients (e.g., Cicchetti et al., 2006), with an emphasis on the limitations of the set of coefficients currently available. Further, two emerging concerns in the assessment of reliability are unitizing reliability assessment, and intracoder reliability assessment, both of which are recommended as independent steps in reliability assessment.

Validity in content analysis

A measure's validity is the extent to which it taps the desired construct. Although validity is as important to content analysis as it is to other methods such as experiments and surveys, validity assessment has been the exception rather than the rule in content analysis.

Some CATA researchers have practiced validation for quite some time (e.g., Gottschalk & Bechtel, 2008). Short, Broberg, Cogliser, and Brigham (2010) have provided an illustration of a comprehensive model for validity assessment of CATA, including a rigorous consideration of construct validation procedures, including tests for content validity, external validity, dimensionality, and predictive validity.

External validity is the extent to which results of a study may be generalized beyond the study sample at hand. Thus, both the careful selection of a population of messages to which the researcher hopes to generalize and the representativeness of the sample taken from that population are of concern to the content analyst.

The integration of content analysis with additional data

In its basic form, a content analysis may be simply descriptive of message content. However, the scope may be productively expanded by combining content analysis data with evidence about the message source, and/or data about the message receiver. Termed an "integrative" approach to content analysis, this collation of content analysis data with additional data offers a means of determining relationships that can help identify the antecedents of message creation such as source factors and the outcomes of the reception of specific messages.

Content analytic data have been collated with source data such as ownership type of organizational sources (such as newspapers), political orientations of source groups, or

demographic, social, and psychographic characteristics of individual sources. Alternatively, data from content analyses have been collated with receiver data such as individual perceptual outcomes and opinions, or aggregate public opinion indicators. An example is Morris's (2009) study of *The Daily Show*'s "Indecision 2004" coverage of the Democratic and Republican party conventions, finding via content analysis harsher disparagement humor during the Republican convention than during the Democratic convention. Panel data collected by the National Annenberg Election Survey during the period found exposure to this convention coverage by *The Daily Show* to be associated with increased negativity toward President Bush and Vice President Dick Cheney.

Qualitative approaches to content analysis

Although quantitative content analysis remains the dominant method in political communication, the use of qualitative content analysis, also known as "ethnographic content analysis," to interpret symbolic construction of social and cultural meanings and emphasis in political messages in documents/texts has been growing. By documents/texts, qualitative scholars mean all written, verbal, and visual content that is retrievable and can be stored (Altheide & Schneider, 2013).

Schreier (2012) identifies the features of qualitative content analysis as interpretive, naturalistic, situational, reflexive, having emergent flexibility, inductive, case-oriented, and putting emphasis on validity (p. 21). The method relies on identifying thematic patterns in a text (i.e., message or set of messages). The themes are not imposed upon the text from outside (e.g., via a theoretically informed coding scheme) or a priori, but they emerge as the researcher undertakes a close reading of a text. Once themes are identified the analyst looks for thematic patterns in the text. Altheide and Schneider (2013) present ethnographic content analysis as a blend of objective content analysis and participant observation that is intended to reveal "how a researcher interacts with documentary materials" (p. 5).

Quantitative and qualitative approaches should be viewed as complementary. However, there are crucial differences in the sampling methodologies — message sampling in quantitative content analysis tends to be probability-based, to satisfy the requirement of external validity. Sampling in qualitative approaches is purposive, and theoretically informed. It is also informed by context, structure, process, and form of political message. For example, political messages in news, blogs, comments, speeches, and advertising are shaped by organizations and by structural constraints of the medium (e.g., newspaper, cable TV, broadcast TV, talk radio, public radio, online news sources, social media, etc.).

Moreover, the premium placed on *discovery* in qualitative analysis indicates an evolving process, allowing for recalibration as a consequence of reflexivity and constant comparison. Thus, the reliability of a qualitative coding scheme/protocol is dependent on the researcher. The validity lies in theoretically informed design and protocols, including clear presentation of the rationale for sampling, unit of analysis, coding scheme, and textual analysis (including constant comparison and contrast).

Finally, a significant part of the explanation for differing research goals in the two approaches is that the quantitative approaches privileges predicting the nature, type, and potential outcomes of political messages in media content, whereas the qualitative approach privileges discovery and affecting social and political change.

In recent years, qualitative content analysis has enjoyed increased options in computer software to handle large data sets (e.g., CAQDAS, Code-A-text, Ethnograph MAXQDA, NVIVO10, QDA Miner, Symphony Content Analysis).

Content analysis applications to political communication

While a vast range of communications might be considered to fall under the rubric of "political," this discussion of the application of content analysis will focus on the traditional meanings of the term, inclusive of formal political processes such as news about elections and elected officials, political editorializing, and political campaigns.

The news function

Analysis of news media content is used to study how the distribution of news sources, gatekeeping, agenda-setting and agenda-building, framing, and priming explain the functions of news in the context of daily politics, elections, and civic engagement in democratic societies. Through systematic examination of what the news media cover, we learn how the gatekeeping function influences what is relevant for political discussion. It is possible to explain what editors block out from the coverage by comparing media coverage with information from other sources. In combination with public opinion surveys, analysis of the topics covered in news allows researchers to show how the news media play a role in setting the political agenda. In comparative studies, systematic analysis of content is used to study similarities and differences and how they relate to differences in media and political systems.

Content analysis is often the method of first choice in framing studies, with scores of studies evident in the literature. From a source perspective, political communication scholars have used content analysis to explain how complex political issues are presented and packaged in accessible forms or as "clusters of messages" (i.e., news frames; Entman, 1993). Whereas, from a receiver perspective, scholars have used the analysis of news content to show how some meanings are preferred over other competing meanings, thereby influencing public opinion and voting behavior (Iyengar, 1991; Semetko & Valkenburg 2000). Analyses of news content in the studies on distribution of news sources and news frames have explained how government officials and authoritative sources get their message out and become primary definers of the political issues. Through a series of studies using content analysis, Lance Bennett and others have shown that news content with political import is "indexed" to statements by official sources and more specifically the governing elites (Bennett, Lawrence, & Livingston, 2006).

Even though large-scale studies have not shown strong evidence of news media bias, content analysis of particular news outlets has shown evidence of bias toward partisan ideology or slant in political coverage to nationalism and political party (Schiffer, 2006).

In recent years, scholars have used content analysis to study how websites, blogs, and online forums foster a sense of community and civic engagement. Content analysis of online forums has shown that some individuals on such forums continue to play influential opinion leadership roles in mediating and amplifying the political messages (e.g., Himelboim, Gleave, & Smith, 2009).

Behavioral studies have used content analysis in combination with surveys and experiments to understand the priming function and effectiveness of messaging via campaigns, news, and political advertisements.

Visual images on television and now increasingly on social media is an area that has not attracted a lot of attention from political communication scholars; however, the utility of content analysis in studying images and audio has been demonstrated compellingly by Grabe and Bucy (2009) in a series of studies. They have developed coding schemes to measure visual framing, camera angles, and duration of shots, confirming the power of visuals in political messaging.

The campaign function: Platforms, speeches, debates, and ads

While most content analyses of political campaign communications have focused on particular communication types, a comprehensive treatment is provided by Benoit's (2007) book on political campaigns. The text includes summary content analyses of ads, debates, brochures, TV talkshow appearances, and Web content, applying his functional theory analysis across this broad range of media used in campaigns.

The formal presentation of political party positions is routinely communicated via platforms and speeches, and these content types have been the focus of dozens of content analyses. There is a long history of applications of CATA techniques to political message content, including key examples of analyses of party platforms and of political speeches. The Manifesto Project has utilized a carefully developed coding protocol for the analysis of party election programs, and have compiled an online database of 3,611 political manifestos from 623 elections across 55 nations (https://manifestoproject.wzb.eu/).

Studies of political speeches have variously examined the verbal content of the speeches, the nonverbal content, and presentation styles of mediated (e.g., televised) speeches. CATA has been applied, as in the case of Bligh and Robinson's (2010) unique application of Diction to archived speeches by Gandhi, demonstrating the prevalence of eight constructs measuring "charismatic" content. Human coding has been used by Waheed, Schuck, de Vreese, and Neijens (2011) to adapt Schwartz's Basic Human Values scale to tap values expressed in political speeches by six female leaders from developed and developing countries, and by others to analyze speeches in other nations.

Although studies focusing on campaign speeches have predominated, others have examined political speeches unrelated to campaigns, including political speech on websites or social media. For example, Druckman, Kifer, and Parkin (2010) studied negativity on over 700 US congressional candidate websites, finding support for the normalization thesis that online behavior mimics that found in traditional media. And Tumasjan, Sprenger, Sandner, and Welpe (2010) used the LIWC program to conduct

a CATA analysis of political sentiment in over 100,000 Twitter posts related to the 2009 German national election.

Although dozens of studies have examined the impact of viewing US presidential debates, fewer have content analyzed those debates. The studies that exist generally take a "horserace" perspective. Bechtolt, Hilyard, and Bybee (1977) concluded after a content analysis of the 1976 US presidential debates between Ford and Carter that Carter "exercised more control" by initiating more topics, and spending more time on those topics. Later researchers have used both CATA and human coding to examine differences among debating candidates in terms of characteristics such as communicator style, both in the United States and elsewhere.

While most studies have analyzed only the verbal content of debates, often relying on transcripts, others have examined the pictorial treatment of the candidates. Tiemens, Hellweg, Kipper, and Phillips (1985) combined verbal and visual measures in their content analysis of the 1980 Carter–Reagan debates, using a coding scheme developed by Tiemens for earlier debates.

Scores of content analyses have looked at political advertising, making it the most studied aspect of campaign politics. Typical motivators for such analyses include persuasion theories (including those that query the role of negative appeals, emotional appeals, and communication style or tone), agenda-setting theory, and cross-cultural comparisons.

Brader (2006) executed a broad-based content analysis of over 1,500 political TV ads, measuring over 100 variables per ad, including negativity appeals, informational appeals, and a host of nonverbal/visual elements denoting emotional tone.

Kaid and Johnston (2001) present their comprehensive "videostyle" analysis, focusing on a political ad's verbal content, nonverbal content, and film/video production techniques (p. 27) (i.e., including both substance and form variables). Their analysis of all TV ads for US presidential candidates from 1952 to 1996, collected from the Political Commercial Archive at the University of Oklahoma, found that the commercials of winners were more likely to use logical appeals, emphasize competency and values, attack the record of an opponent, use testimonials rather than anonymous announcers, use more special effects, and be of shorter length.

How political advertising may help set the news media agenda has been a focus of a number of studies. For example, Lopez-Escobar, Llamas, McCombs, and Lennon (1998), in a study of the 1995 Spanish elections, showed that newspaper and TV political ads influenced news coverage, and vice versa.

Future horizons

Political discourse takes place via an increasing array of media platforms, such as blogs, microblogs (e.g., Twitter), social media, and on-demand moving image media. The diversification of content and delivery modes poses ever greater potential for content analysis to be utilized and to be integrated with other methods to ratchet up its descriptive and predictive power in the arena of political communication.

At the center of research in political communication are spoken and written words as primary sources of data to understand, explain, and predict political effects of actions of states, political parties, activist groups, and more often individuals. In the past, the cost and logistical difficulties of collecting large message data sets was a constraint for content analysts. However, with digital technologies and the Web, collecting and analyzing very large data sets, also known as "big data," has become practical. In an era of "big data" discovery and analysis, the role of content analysis is very "big" indeed. The potential for linking political content analytic measures such as online posts, public opinions, and social media "likes" with behavioral indicators such as voting intentions has become much greater. The availability of large data sets of text suggests that use of content analysis will only become more widespread in all disciplines of the social sciences, but notably so in political communication.

SEE ALSO: Advertising, Political; Agenda-Setting; Blogging, Political; Civic Engagement; Election Campaign Communication; Emotion in Politics; Framing Theory; Gatekeeping; Humor & Satire, Political; Ideology; Image, Political; Indexing Theory; Internet; Journalism, Political; Language, Political; Mass Communication; Media Bias; Media Ethnography; News Values; Online News; Opinion Leadership; Partisanship; Persuasion, Political; Political Communication; Political Communication Research; Political Communication Research Methods; Political Discourse; Priming; Print Media; Public Opinion Polls; Rhetoric, Political; Social Media; Talkshows, Political; Television; Values; Visual Communication

References

Altheide, D. L., & Schneider, C. J. (2013). *Qualitative media analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Bechtolt, W. E., Hilyard, J., & Bybee, C. R. (1977). Agenda control in the 1976 debates: A content analysis. *Journalism Quarterly, 54,* 674–681, 689.

Bennett, L., Lawrence, R. G., & Livingston, S. (2006). None dare call it torture: Indexing and limits of press independence in the Abu Ghraib scandal. *Journal of Communication, 56,* 467–485.

Benoit, W. L. (2007). *Communication in political campaigns.* New York, NY: Peter Lang.

Bligh, M. C., & Robinson, J. L. (2010). Was Gandhi "charismatic"? Exploring the rhetorical leadership of Mahatma Gandhi. *The Leadership Quarterly, 21,* 844–855.

Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work.* Chicago, IL: University of Chicago Press.

Cicchetti, D., Bronen, R., Spencer, S., et al. (2006). Rating scales, scales of measurement, issue of reliability: Resolving some critical issues for clinicians and researcher. *Journal of Nervous and Mental Disease, 194,* 557–564.

Druckman, J. N., Kifer, M. J., & Parkin, M. (2010). Timeless strategy meets new medium: going negative on congressional campaign web sites, 2002–2006. *Political Communication, 27,* 88–103.

Entman, R. M. (1993). Framing: Towards a clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.

Gottschalk, L. A., & Bechtel, R. J. (Eds.). (2008). *Computerized content analysis of speech and verbal texts and its many applications.* New York, NY: Nova Science Publishers.

Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections.* Oxford, UK: Oxford University Press.

Himelboim, I., Gleave, E., & Smith, M. (2009). Discussion catalysts in online political discussions: Content importers and conversation starters. *Journal of Computer-Mediated Communication, 14,* 771–789.

Iyengar, S. (1991). *Is anyone responsible? How television frames political issues.* Chicago, IL: University of Chicago Press.

Kaid, L. L., & Johnston, A. (2001). *Videostyle and content of televised political advertising.* Westport, CT: Praeger.

Liu, B. (2012). *Sentiment analysis and opinion mining.* San Rafael, CA: Morgan & Claypool.

Lopez-Escobar, E., Llamas, J. P., McCombs, M., & Lennon, F. R. (1998). Two levels of agenda setting among advertising and news in the 1995 Spanish elections. *Political Communication, 15,* 225–238.

Morris, J. S. (2009). *The Daily Show with Jon Stewart* and audience attitude change during the 2004 party conventions. *Political Behavior, 31,* 79–102.

Pennebaker, J. W., & Chung, C. K. (2008). Computerized text analysis of al-Qaeda statements. In K. H. Krippendorff & M. A. Bock (Eds.), *A content analysis reader* (pp. 453–466). Thousand Oaks, CA: Sage.

Schiffer, A. (2006). Assessing partisan bias in political news: The case(s) of local Senate election coverage. *Political Communication, 23,* 23–39.

Schreier, M. (2012). *Qualitative content analysis in practice.* Los Angeles, CA: Sage.

Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication, 50*(2), 93–109.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods, 13,* 320–347.

Tiemens, R. K., Hellweg, S. A., Kipper, P., & Phillips, S. L. (1985). An integrative verbal and visual analysis of the Carter–Reagan debate. *Communication Quarterly, 33,* 34–42.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.* Menlo Park, CA: AAAI Press.

Waheed, M., Schuck, A., de Vreese, C., & Neijens, P. (2011). More different than similar: Values in political speeches of leaders from developed and developing countries. *Journal of Intercultural Communication, 26.*

Further reading

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

Riffe, D., Lacy, S., & Fico, F. (2005). *Analyzing media messages: Using quantitative content analysis in research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.

Kimberly A. Neuendorf (PhD, Michigan State University) is professor of communication in the School of Communication at Cleveland State University, USA. She is the author or coauthor of over 80 peer-reviewed research articles. Her research interests

include media and marginalized populations, the analysis of spectator responses to the moving image, the adoption of new communication technologies, and the method of quantitative content analysis. The second edition of her text *The Content Analysis Guidebook* (2002) is set to appear in 2015.

Anup Kumar is associate professor of communication in the School of Communication, Cleveland State University, USA. He started as an environmental scientist and then made a transition into journalism. He was a practicing journalist for many years, after which he made a second career transition into academia. He is the author of *The Making of a Small State: Populist Social Mobilisation and the Hindi Press in the Uttarakhand Movement* (2011) and research articles on the interaction between the news media and politics.