Online Research Methods, Qualitative

1 author:

Lynne Webb
Florida International University
97 PUBLICATIONS   **679** CITATIONS

Online Research Methods, Qualitative

LYNNE M. WEBB
Florida International University, USA

A researcher might examine messages from new media whenever s/he desires to study human communication by examining the textual residue of human interaction. For example, political blogs offer glimpses into debates of contemporary, important, and complex issues such as deliberations during national elections. Alternatively, an examination of social media websites offers insight into online behavior per se, such as self-disclosure (e.g., Tang & Wang, 2012), social presence (Luzon, 2011), self-presentation strategies (e.g., Sanderson, 2008), negotiation of social roles (Suter, Baxter, Seurer, & Thomas, 2014), and displays of gender identity (e.g., Willem, Arauna, Crescenzi, & Tortajada, 2010). New media researchers conduct primarily three types of studies:

- Researchers interview or survey Internet users to discover their perceptions and motivations.
- Researchers examine the design characteristics of websites per se by assessing elements such as the linking patterns and documenting connections between such elements and outcome variables such as popularity.
- Researchers analyze website content typically to achieve one of two goals: to contrast new media to traditional media or to examine important characteristics of posts, such as the language characteristics.

Given that (i) other entries in this volume address surveys and interviewing, and (ii) studies of design characteristics tend to employ quantitative methods, the focus of this entry will be on the third element, namely, examinations of website content. Researchers who study website content examine text in either written or visual form (pictures and videos) or occasionally in both forms simultaneously; hereafter the term "text" is used to describe both written and visual content.

Online content provides low cost (or no cost) data pools that are publicly and readily available for downloading and analyzing. Webb and Wang (2013a) identify several advantages to online qualitative methods versus traditional research methods in communication.

- With no initial investment of capital for equipment, researchers can simply harvest datasets.
- Because the data preexists in the public domain, researchers can abandon many labor-intensive activities such as training confederates.

- Researchers can collect data as quickly as they can download text; data collection becomes almost instantaneous.
- Recently downloaded data are "fresh, contemporary, and immediate—allowing timely research of important, contemporary developments" (p. 210)
- The researcher can download the exact and complete dataset with no missing data or nonresponse issues.
- The interactivity of social media posters and commenters affords the opportunity to study "public response to opinions, including which issues are discussed and which are not, the valence of responses, and the strength of responses as evidenced both in language as well as number of repostings." (p. 210)
- Finally, user interconnectivity allows researchers to follow trails of influence across adoptions and innovations with documented evidence of changes.

Selecting the object of study

When researchers study online content, they must select which text to analyze. Given that websites can contain thousands of posts and comments across multiple topics and many years, researchers select text based upon the focus of the given study, specifically research questions and hypotheses. However, the vast majority of new media researchers who examine text draw samples from one of three objects of study:

- entries—original website posts such as Facebook status updates,
- comments—responses to original entries or to previous comments, or
- conversational strings or threads—posts and comments on a given topic, such as a blog entry and relevant comments to that entry.

Each of these three options allows for the examination of an interesting object of study. Researchers who focus on entries examine the discourse of would-be opinion leaders. When examining comments, researchers can look at decoding, sense-making, information processing, and additional communication-related activities that lie at the heart of influence and persuasion. A focus on strings/threads allows observation of users expressing agreement and disagreement, reaching consensus, and/or engaging in conflict. When websites contain archival posts, across-time data can reveal how the number of entries, comments, and strings increase or decrease on a given topic in response to online discourse and/or current events. Alternatively, researchers may examine posts across multiple websites on a given topic during a fixed period of time.

Each type of new media presents written and/or visual text for analysis; indeed, communication scholars have examined blogs, microblogs, discussion-board entries, email messages, Facebook posts and comments, search engine output, social media profiles, summary reports written by online project teams and posted online, transcripts of online interviews, as well as YouTube videos. Furthermore, an increasing number of researchers examine content from multiple online media channels within the same dataset. For example, Gilpin (2010) examined one company's online press releases,

blog posts, and Twitter messages to gain a more comprehensive understanding of an organization's online presence.

Sampling the text for analysis

After selecting the object of study, the researcher must select the exact text for analysis. For example, researchers who study LinkedIn posts must decide which of the millions of existing posts will be selected for analysis. The selection of exact text from all available text of its type is called sampling. Internet researchers employ conventional sampling techniques in an attempt to obtain samples that accurately represent the diversity in the population under study. An advantage to studying online messages is that the researcher can define the object of study so narrowly (e.g., campaign websites of all declared candidates in a primary election) that a complete census of the population of messages can be obtained and studied. If the population is larger (all Facebook user profiles from a given Facebook network displaying rainbow colors following the U.S. Supreme Court ruling on gay marriage), random sampling can be employed. Additionally, random assignment can be used across the data collection and analysis processes from random assignment of team members to specific sections of the dataset for downloading to random assignment of coders to subsets of the sample for coding (Webb, Gibson, Wang, Chang, & Thompson-Hayes, 2015). Webb and Wang (2013b) offer detailed descriptions of how to apply multiple sampling techniques to data scraped from the Internet.

Types of qualitative analysis

Many qualitative techniques are largely absent from the study of online messages including focus groups, oral histories, and analyses of historical documents (e.g., letters and diaries). Instead, qualitative studies of new media are field studies of naturally occurring messages and conversations. Researchers examine the messages and interactions per se via various forms of textual analysis as well as analyze the same text for deeper meanings such as examining discourse on fan websites to understand how message receivers decode their object of fandom (Webb, Chang, Hayes, Smith, & Gibson, 2012). Studies of online text rarely employ innovative data analytic methods and instead rely almost exclusively on a limited number of traditional qualitative techniques listed below. Using these techniques, researchers can analyze the same dataset multiple ways and produce multiple research reports from the same dataset.

- online interviews (e.g., Martey & Consalvo, 2011);
- participant-observations often reported using autoethnographic techniques (e.g., Honeycutt, 2005);
- case studies (e.g., Halse, 2014);
- content analyses (e.g., Willem et al., 2010);

- thematic analysis using a grounded theory approach (e.g., McNeil, Brna, & Gordon, 2012);
- discourse analysis examining language per se (e.g., Florini, 2014); and
- framework analysis (Fotiadou, Northcott, Chatzidaki, & Hilari, 2014).

Online interviews follow the conventions of face-to-face interviewing but occur online via video-chat, texting tools, or the unique affordances of specific online media. For example, Martey and Consalvo (2011) interviewed 30 online gamers via voice and text chat in *Second Life*. Researchers typically analyze the transcripts of such interviews as text using one of the four techniques: content analysis, discourse analysis, framework analysis, or thematic analysis.

Participant-observations. Researchers can participate in social media and online groups with the intention of reporting their observations and experiences, often using ethnographic or autoethnographic techniques to report their experiences. For example, Honeycutt (2005) provides a detailed description of a verbally abusive hazing ritual for newcomers to one online community.

Case studies examine only a few examples of a phenomenon in depth to identify central concerns. Researchers employ case studies on two occasions: to closely examine rare phenomena and to begin a line of inquiry with an in-depth, exploratory study to gain preliminary insights. Researchers employ three guiding principles to select cases: typicality, complexity, and conformity.

- Typicality. Researchers often select cases that they believe are typical of the phenomenon under study. They may randomly sample cases to increase the probability of selecting typical cases and to avoid introducing bias in case selection. For example, Luzon (2011) randomly selected 11 academic blogs from weblog directories to examine social and antisocial behavior on academic blogs. Exploratory research often uses typical cases.
- Complexity. In contrast, researchers may intentionally examine particularly complicated cases to reveal the multiple issues at play in complex phenomenon. For example, Halse (2014) examined comments on only two fan discussion boards (one from the United States and one from Norway) surrounding only one character from one television program. However, the character defied a common ethnic stereotype, prompting contested discussion among fans in two cultures, and thus allowing perceptions of ethnic stereotypes to emerge. Complicated cases are not necessarily typical but they can provide insight into important issues.
- Conformity. When the researcher begins with hypotheses or very specific research questions, s/he may select cases that conform to the criteria necessary to test the hypotheses thoroughly or to answer the research questions definitively. For example, Trammel (2006) tested the Functional Theory of Political Campaign Discourse by examining "attacks" on presidential candidates' blogs.

Content analysis. When conducting a content analysis, researchers begin with preselected categories and then examine the data to discover to what extent the data fit the categories. For example, Willem et al. (2010) analyzed anonymous pictures on a popular

Spanish social networking website using Goffmanian categories of gender display and performance. Researchers preselect categories based upon (i) to test a theory/principle, or (ii) to ensure consistency in the coding of characteristics under study. For example, comments may agree with the original post, disagree, both agree and disagree, as well as be irrelevant.

Software programs are available to search for keywords and valences, but most communication researchers report coding "by hand," even for complex, multilevel analyses. Researchers often determine coder agreement by assigning each coder the same small set of data to code independently and then compare results. If overlapping data is coded at the beginning of the project, then disagreements can be settled prior to the coding of the bulk of the data. Often coders continue to independently code small portions of data and compare outcomes until they reach high agreement and then code the bulk of the dataset. Coders can settle disagreement by devising decision rules that become part of the codebook. Such a preliminary process is often described as training in the research report and such training increases the odds of obtaining high intercoder reliability scores. "High agreement across multiple coders offers evidence of reliability of the analysis and thus increases the credibility of the results. Most published thematic analyses report intercoder agreement at 88% or higher but reviewers prefer over 95% agreement" (Webb & Wang, 2013a, p. 217).

Thematic analysis. Thematic analysis, also known as grounded theory analysis, is among the most common techniques for textual analysis in communication scholarship. Furthermore, many communication scholars have employed grounded theory and/or thematic analysis "including research in the traditions of positivist, interpretive, modern, post-modern, feminist, and critical scholarships" (Gibson & Webb, 2012, p. 164). Thematic analysis involves open coding to identify themes that emerge from the data rather than the researcher imposing a priori categories on the data (as is the case in the content analytic method described earlier). During open coding, researchers identify commonly occurring ideas (called themes) that reoccur in the dataset. For example, McNeil et al. (2012) discovered that Twitter posts containing the word "seizure" fall into four dominant categories or themes: metaphorical, personal accounts, informative, and ridicule/joke. Thematic analysis allows the voices of the text creators, often referenced as informants, to emerge from their accounts; such analyses offer scholars insight into the thinking and lived experience of the Internet users who created the textual accounts under analysis and who have first-hand experience with the phenomenon under study.

Communication researchers often employ Owen's (1984) three criteria of a theme:

- recurrence (the same *idea* reappears across the dataset, although the words, manner, and style of expression may vary);
- repetition (the same or nearly the *same words* express the same idea across the dataset); and
- forcefulness (*ideas strongly stressed* verbally or nonverbally via dramatic language, vivid imagery, or textual emphasis such as italicizing a passage).

Researchers examining smaller samples (20 or fewer sources) may define a theme as two or more sources within the dataset expressing the same idea (e.g., Instagram posts from two users display pictures of the same soccer score). However, researchers usually require three or more sources. Major themes emerge when almost all the users in the dataset mention the same idea. To locate themes, coders read the data, listing themes as they emerge; then the coders reread the data set using a constant comparison process until they reach theoretical saturation (i.e., no new themes emerge during a read-through or when another case is added to the dataset).

Gibson and Webb (2012) identified multiple advantages and applications of analytic techniques that allow themes to emerge from the data:

- Human communication behavior is often governed by covert, unwritten, unvoiced understandings. Thematic analysis allows users' understandings of covert and unofficial rules, norms, and patterns of online behavior to emerge.
- Internet users' naïve theories can emerge, thus allowing the researcher insight into how users think about and engage in their online interactions.
- The users' authentic experiences and thinking emerge, thus offering fresh perspectives and insights into their online communication.
- Many aspects of online communication remain unexplored. Thematic analysis offers the necessary tools to provide an initial understanding of these virgin territories.
- Emergent themes may identify key concepts at play in the phenomenon under study and thus lay the groundwork for the development of a questionnaire later used in quantitative research.
- There are fewer theories about online communication versus face-to-face communication. Thematic analysis allows for the emergence of new, grounded theories to describe and explain online communication.
- Thematic analysis allows for users' complex understandings to emerge that may provide clarity to and integration of existing theories.

Discourse analysis. Like thematic analysts, discourse analysts examine text to discover patterns but focus on language, language use, and/or linguistic qualities, such as the use of emoticons. Discourse analysts employ the same methodological procedures as thematic analysts to study linguistic objects of study. For example, Florini (2014) identified linguist practices (figurative language, indirectness, double-ness, and wordplay) used by Twitter users to display African American identity. Perhaps the most common form of discourse analysis, metaphor analysis, provides researchers insights into users' thinking by examining the metaphors employed in online text.

Framework analysis offers a middle ground between content analysis and thematic analysis as researchers begin with broad categories but search for emergent themes within those categories. For example, Fotiadou et al. (2014) examined blogs written by users with aphasia to discover ways aphasia impacts (i) family relationships, (ii) additional social relationships, and (iii) bloggers' sense of self. Hypothesized categories often reflect the theoretical or conceptual underpinnings of the study. Such categories

allow researchers to search for themes within rather than necessarily across categories, thus adding meaningful limitations to the amount of data needed to reach saturation and end the constant comparison process. Framework analysis is used primarily to examine changes in thinking and the development of innovations, often about public policy or health-care issues. The software NVivo offers options for framework analysis with large datasets.

Computer analysis. Qualitative research is notoriously labor-intensive especially with large datasets scraped from the Internet; Gibson and Webb (2012) offer a detailed description of that four-step, labor-intensive process (open coding, axial coding, selective coding, and emergence of theory). Qualitative researchers increasingly employ computer software to address this concern. A few software programs are designed specifically to analyze text scraped from the Internet; see Boulos, Sanfilippo, Corley, and Wheeler (2010) for a review of such programs. Other software programs can be used to analyze any text-based dataset including the analysis of online text. Data analytic programs range from the simple to the complex and ultimately allow analysis on multiple levels simultaneously.

SEE ALSO: Autoethnography; Axial Coding; Case Study Research; Code List/ Codebook; Constant Comparison; Data Saturation; Discourse Analysis; Frame Analysis; Grounded Theory; Interviewing Vulnerable Populations; Interviewing, Qualitative; Metaphor Analysis; Online Ethnography; Qualitative Analysis Software (ATLAS.ti/Ethnograph/MAXQDA/NVivo); Representation/Representativeness; Sampling, Qualitative; Thematic Analysis/Interpretive Thematic Analysis

References

Boulos, M. N. K., Sanfilippo, A. P., Corley, C. D., & Wheeler, S. (2010). Social web mining and exploitation for serious applications: Technosocial predictive analytics and related technologies for public health, environmental and national security surveillance. *Computer Methods and Programs in Biomedicine*, *100*, 16–23. doi:10.1016/j.cmpb.2010.02.007

Florini, S. (2014). Tweets, tweeps, and signifyin': Communication and cultural performance on "Black Twitter." *Television and New Media*, *15*, 223–237. doi:10.1177/1527476413480247

Fotiadou, D., Northcott, S., Chatzidaki, A., & Hilari, K. (2014). Aphasia blog talk: How does stroke and aphasia affect a person's social relationships? *Aphasiology*, *28*, 1281–1300. doi:10.1080/02687038.2014.928664

Gibson, D. M., & Webb, L. M. (2012). Grounded-theory approaches to research on virtual work: A brief primer. In S. D. Long (Ed.), *Virtual work and human interaction research: Qualitative and quantitative approaches* (pp. 160–175). Hershey, PA: IGI Global Publishers.

Gilpin, D. (2010). Organizational image construction in a fragmented online media environment. *Journal of Public Relations Research*, *22*, 265–287. doi:10.1080/10627261003614393

Halse, R. (2014). Textual poaching, gamekeeping and the counter-culture-stereotype. *Nordicom Review*, *35*, 65–79. doi:10.2478/nor-2014-0005

Honeycutt, C. (2005). Hazing as a process of boundary maintenance in an on-line community. *Journal of Computer-Mediated Communication*, *10*(2). doi:10.1111/j.1083-6101.2005.tb00240.x

Luzon, M. J. (2011). "Interesting post, but I disagree": Social presence and antisocial behaviour in academic weblogs. *Applied Linguistics*, *32*(5), 517–540. doi:10.1093/applin/amr021

Martey, R. M., & Consalvo, M. (2011). Performing the looking-glass self: Avatar appearance and group identity in *Second Life*. *Popular Communication*, *9*, 165–180. doi:10.1080/15405702.2011.583830

McNeil, K., Brna, P. M., & Gordon, K. E. (2012). Epilepsy in the Twitter era: A need to re-tweet the way we think about seizures. *Epilepsy & Behavior*, *23*(2), 127–130. doi:10.1016/j.yebeh.2011.10.020

Owen, W. F. (1984). Interpretive themes in relational communication. *Quarterly Journal of Speech*, *70*, 274–287. doi:10.1080/00335638409383697

Sanderson, J. (2008). The blog is serving its purpose: Self-presentation strategies on 38pitches.com. *Journal of Computer-Mediated Communication*, *13*, 912–936. doi:10.1111/j.1083-6101.2008.00424.x

Suter, E. A., Baxter, L. A., Seurer, L. M., & Thomas, L. J. (2014). Discursive constructions of the meaning of "family" in online narratives of foster adoptive parents. *Communication Monographs*, *81*, 59–78. doi:10.1080/03637751.2014.880791

Tang, J. H., & Wang, C. C. (2012). Self-disclosure among bloggers: Re-examining of social penetration theory. *Cyberpsychology, Behavior, and Social Networking*, *15*(5), 245–250. doi:10.1089/cyber.2011.0403

Trammel, K. D. (2006). Blog offensive: An exploratory analysis of attacks published on campaign blog posts from a political public relations perspective. *Public Relations Review*, *32*, 402–406. doi:10.1016/j.pubrev.2006.09.008

Webb, L. M., Chang, H. C., Hayes, M. T., Smith, M. M., & Gibson, D. M. (2012). How fans make sense of *Mad Men*: An analysis of commentary from online fan websites. In J. C. Dunn, J. Manning, & D. Stern (Eds.), *Lucky strikes and a three-martini lunch: Thinking about television's Mad Men* (pp. 224–236). Newcastle, UK: Cambridge Scholars Publishing.

Webb, L. M., Gibson, D. M., Wang, Y. X., Chang, H. C., & Thompson-Hayes, M. (2015). Selecting, scraping, and sampling big data sets from the Internet: Fan blogs as exemplar. *SAGE Research Methods Cases* [19 pages]. doi:10.4135/9781446727305014558403

Webb, L. M., & Wang, Y. X. (2013a). Techniques for analyzing blogs and micro-blogs. In N. Sappleton (Ed.), *Advancing research methods with new technologies* (pp. 183–204). Hershey, PA: IGI Global Publishers. doi:10.4018/978-1-4666-3918-8.ch012

Webb, L. M., & Wang, Y. X. (2013b). Techniques for sampling on-line data sets. In W. C. Hu & N. Kaabouch (Eds.), *Big data management, technologies, and applications* (pp. 95–114). Hershey, PA: IGI Global Publishers. doi:10.4018/978-1-4666-3918-8.ch012

Willem, C., Arauna, N., Crescenzi, L., & Tortajada, I. (2010). Girls on Fotolog: Reproductions of gender stereotypes or identity play? *Interactions: Studies in Communication & Culture*, *2*, 225–242. doi:10.1386/iscc.2.3.225_1

Further reading

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics*. Hoboken, NJ: John Wiley & Sons.

Hesse-Biber, S. N., & Crofts, C. (2010). User-centered perspectives on qualitative data analysis software. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 655–674). New York: Guilford Press.

Hine, C. (Ed.) (2005) *Virtual methods: Issues in social research on the Internet*. Oxford: Berg.

Janert, P. K. (2010). *Data analysis with open source tools*. Sebastopol, CA: O'Reilly Media.

Long, S. D. (2012). *Virtual work and human interaction research: Qualitative and quantitative approaches*. Hershey, PA: IGI Global Publishers.

Russell, M. A. (2011). *Mining the social web: Analyzing data from Facebook, Twitter, LinkedIn, and other social media sites*. Sebastopol, CA: O'Reilly Media.

Lynne M. Webb (PhD, University of Oregon, 1980) is professor in communication, Florida International University. Her work on interpersonal communication and social media has appeared in numerous national and international journals including *Computers in Human Behavior* and the *International Journal of Social Research and Methodology* as well as prestigious edited volumes including *Advancing Research Methods with New Technologies* (IGI Global, 2013), *Families in a Digital Age* (Peter Lang, 2015), and the *Handbook of Research on the Social Impact of Digital Media* (IGI Global, 2015). She has coedited three scholarly readers including *Computer-Mediated Communication in Personal Relationships* (Peter Lang, 2011).