Content Analysis

Chapter Summary

Introduction

This chapter deals with *content analysis*, an analysis technique used to study texts and media objects. The method studies content of media communications in both qualitative and quantitative ways, and includes analysis of words, themes, and value positions. The related methods include *semiotics*, a science of signs in communication, and *hermeneutics*, a science of text interpretation. Two other methods of content analysis relate to the study of language: these are *conversation analysis* (structural analysis of a conversation), and *discourse analysis* (how the view of the world is produced in a discourse, or how the power relations are reproduced through a discourse). *Critical discourse analysis* examines the political nature of the examined texts, considering the issues of power hierarchies, structural inequalities, and historical political struggles. The chapter concludes by considering advantages and drawback of content analysis.

Personal Documents

Diaries, Letters, and Autobiographies

In evaluating personal documents, the *authenticity* criterion is important. Is the purported author of the letter or diary the real author? In terms of of *credibility*, Scott (1990) observed that there are at least two major concerns with respect to personal documents: their factual accuracy and whether they express the true thoughts and feelings of the writer. Scott recommended a healthy skepticism regarding the sincerity with which the writer reports his or her feelings. *Representativeness* is a concern in assessing these materials. Low literacy rates in the past meant that letters, diaries, and autobiographies were preserved by a small class of wealthy, literate people, usually men. A further problem is the selective survival of documents like letters—many documents are damaged, lost, or thrown away. Finally, it is difficult to understand the meaning of the documents we do, since the writer may have used abbreviations or codes that are difficult to decipher.

Visual Objects

Photographs reveal important information about families. Scott distinguished three types of home photograph: *idealization* (a formal portrait of a wedding party or the family in its finery); *natural portrayal* (an informal snapshot capturing action as it happens, though there may be a contrived component); and *demystification* (depicting the subject in an atypical—often embarrassing—situation). Researchers need to be aware of these different types in order to avoid being deceived by the superficial appearance of images. A particular problem for the analyst of photographs, according to Scott, is judging *representativeness*. Photos that survive the passage of time are very

unlikely to be representative for the simple reason that somebody at some time decided they should be preserved. Another problem relates to the issue of what is *not* photographed.

Government Documents

The state produces a great deal of quantitative statistical information, as well as a great deal of textual material of potential interest. In terms of Scott's (1990) four criteria, government materials can certainly be seen as authentic and as having meaning (in that they are clear and comprehensible to the researcher). However, the credibility criterion requires us to consider whether the documentary source is biased. In fact, some documents can be interesting precisely *because* of the bias they reveal, which suggests that caution is necessary in attempting to treat them as depictions of reality.

Official Documents from Private Sources

Companies (and organizations generally) produce many documents, some of which are in the public domain (e.g., annual reports, press releases, advertisements, public relations material) and some of which is private (e.g., company newsletters, organizational charts, minutes of meetings, memos, correspondence manuals for new recruits). Scott's four criteria reveal private documents to be authentic and meaningful (in the sense of being clear and comprehensible to the researcher); however, issues of credibility and representativeness are still likely to require scrutiny.

Mass Media Outlets

Newspapers, magazines, television programs, films, and other mass media are potential sources for social scientific analysis. Authenticity is sometimes difficult to ascertain in the case of mass media outputs because the authors are not always identified, so it is sometimes difficult to know whether a given account was prepared by someone in a position to know all the facts. Credibility is frequently an issue— it is often the uncovering of error or distortion that is the objective of the analysis. Representativeness may not be an issue with newspaper or magazine articles, since many publications make a point of maintaining a consistent tone or ideological bent. Finally, although the literal meaning of mass media outputs is often clear, it usually takes some reflection and theoretical analysis to appreciate the broader societal impact these forms of communication can have.

Virtual Outputs and the Internet as Objects of Analysis

The vastness of the Internet and its growing accessibility make it a valuable source of documents for both quantitative and qualitative data analysis, but it is important to keep Scott's criteria in mind. First, authenticity: anyone can set up a website, so you have no guarantee that a person offering information (such as financial advice) is an authority. Second, credibility: is the information on the site credible, or might it be distorted for some reason? For example, a site that encourages you to buy stock held by its author might exaggerate its value. Third, given the constant flux of the Internet, it's doubtful that one can ever know how representative websites on a certain topic are. Finally, websites are notorious for a kind of "webspeak" that makes it difficult to comprehend what is being said without some insider knowledge. You should be no less skeptical about websites than about any other kind of document.

What Things Need to Be Analyzed?

Depending on the origins and forms of the material, several things can be counted in content analysis: words, themes, or value positions:

- **Words:** How often does a particular word show up, or how are certain words paired together (e.g., feminist/spirituality, race/politics, rave/drugs)? The counting of words can

demonstrate emphasis, the style of writing, or particular interpretation of events. Similarly, the frequent pairing of words (e.g., why are raves often mentioned together with drugs?) may produce a certain interpretation that leads the reader to see the studied phenomenon in a particular light. Hence uncovering how the words are paired in the text can be a beginning or a more in-depth analysis of the issue.

- **Subjects and themes:** Looking for themes is an example of interpretative approach in content analysis. This includes both *manifest* and *latent* content. **Manifest content** describes what is being said directly, while **latent content** states what is being implied. For example, in the subject of women in politics, looking for themes will involve looking for explicitly stated points on the topic, but it will also describe the themes or stereotypes that are being implied in the text.

- **Value positions:** Researchers may also look for whether the writer of the text adopted a certain value position. Is the subject being presented in a positive or negative light? Is the writer sympathetic to the main actors of the story he or she is presenting? When describing a criminal, is the author putting all the blame on the accused, or does he look at the social conditions of the accused and therefore places less blame on the accused person? If no manifest value position of the writer is stated, can their position be deduced from the latent content? Another way to reveal value positions is by coding the ideologies, beliefs, or principles expressed in the analyzed document.

Coding

There are two key documents that a researcher must develop for coding. There is a *coding schedule*, which is the actual form where the data are recorded, and a *coding manual*, which provides a set of instructions to coders.

Coding Schedule

The *coding schedule* is similar to a spreadsheet. It has the categories listed across the top and one row for each media item being coded, and the codes for each case are recorded directly into the coding schedule. Writing information into a coding schedule facilitates the data transfer to a computer program (e.g., SPSS).

Coding Manual

The coding manual will list things like what is to be coded, the categories subsumed under each dimension, the numbers (codes) that correspond to each category, and general guidance and instructions for coders. The coding manual will describe not only what information about the case is to be coded, but will also propose a code for how the case is described in the analyzed text (position of the author), and a code for describing the location of article in the analyzed journal, newspaper, or the media (front page, editorial, back of the issue, etc.). The categories specified in the coding manual must be sufficiently distinct for coders to be able to code consistently and for the analysis to be meaningful.

Potential Pitfalls in Devising Coding Schemes

Developing a coding scheme is not an easy task and researchers may encounter several problems, which must be addressed early in the process:

- *Mutually exclusive categories*: The coded categories must be mutually exclusive, to avoid coder's confusion about the category where the item can be placed.
- *Exhaustive*: Every possible dimension should have a code that can be attached to it.

- *Clear instructions*: The instructions should be clear enough to prevent discretion on the part of the coders in assigning codes. This means that often the instructions should be pretty elaborate to avoid ambiguity in coding.
- *A clear unit of analysis*: The unit of analysis, or what is being analyzed, should not be confused with the incidence, or the case, being recorded. The **unit of analysis** is the actual entity to be analyzed (e.g., a newspaper article, a TV series episode, a document), while the case being recorded is an instance of a studied problem (e.g., representation of masculinity in police officers' activities).

To check the quality of the coding scheme, a pilot coding of small number of cases should be conducted. Such pilot coding can help identify problems in the coding, or reveal a category that should be coded but is not indicated in the scheme, or a category which, as a result of coding, includes a large number of items. If the latter occurs, the coding category should be broken down into several more specific ones to better represent the variety of cases.

Reliability is also a concern when developing a coding scheme, and pre-testing the scheme can help researchers develop better instructions for coders and as a result improve reliability of the study. This is particularly important when there are several coders on the project and their coding should be consistent to ensure inter-coder reliability.

Content Analysis without a Pre-Existing Coding Scheme

Qualitative content analysis looks for underlying themes that are present in a unit of analysis. The researcher decides what themes are to be extracted and how the presence of a theme is established. Themes are recurrent topics mentioned in a text or in a recording. The researcher extracts the themes based on her previously-developed research plan, or codifying the issues that were previously unanticipated but are recurring in the document.

Ethnographic content analysis (ECA) is one type of qualitative content analysis. It differs from quantitative content analysis in that the researcher is constantly revising the themes or categories obtained from the data. It is a more interactive approach to collecting data, coding, and interpreting. Ethnographic content analysis involves "constant discovery and comparison of relevant situations" (Altheide, 1996), and allows for constant refinement of categories and generation of the new ones. By contrast, categories in quantitative content analysis are predetermined and do not change during the data analysis. ECA also looks at the context in which documents are created. For example, ECA of newspaper articles requires some understanding of how newspapers work and are organized.

Semiotics

Semiotics is the "science of signs," and involves analysis of signs and symbols that occur in everyday life and are part of communicative behaviour. In other words, semiotics studies systems of communication and meaning by looking at signs. Its key terms are the following:
- *The sign* is something that has a meaning and stands for something else, so that the observer can decode a sign and interpret it. (e.g., a red hand pedestrian traffic signal light is a sign that means that the pedestrian should refrain from crossing the street and wait. The sign has two components: a *signifier* and a *signified*.)
- *Signifier* is the thing that indicates an underlying meaning (e.g., the traffic signal light itself)
- *Signified* is the meaning that is pointed to (e.g., "stop and do not cross" meaning conveyed by the red hand traffic signal)
- *Denotative meaning* is the obvious meaning of a signifier, a function it performs (e.g., the denotative meaning of the signal light is that it regulates pedestrian traffic)
- *Connotative meaning* is the secondary meaning that arises with the denotative meaning (e.g., "speed up crossing the street to avoid the red light")

- *Polysemy* recognizes that signs may be interpreted in many different ways.

Semiotics works to uncover hidden meanings in text or messages, and its main strength is in looking beyond the ordinary and the everyday. However, semiotic interpretations depend on the person doing the analysis, and his or her interpretation of the discovered meanings can be somewhat arbitrary.

Hermeneutics

Hermeneutics is the science of text interpretation, especially the interpretation of biblical and other religious texts. It was initially applied to the interpretation of scripture. Hermeneutics consists of theory and methodology of text interpretation, and is the process of looking at the whole of a text from within the context of its parts, and looking at the parts of the text from the broader context of the whole. Its main idea is that deriving meaning of a text must be made from the perspective of the text author, and must take into account the contextual circumstances in which the text was constructed.

Two Approaches to the Study of Language

Conversation analysis and discourse analysis represent two methods of studying language. Both include quantitative and qualitative aspects.

Conversation Analysis

Conversation analysis (CA) examines how social order is created through communication. It derives from ethnomethodology, which studies ways in which people negotiate everyday situations by focusing on common-sense reasoning. Everyday communication and use of practical reasoning helps people to establish taken-for-granted patterns of interaction through which social order is recreated. Hence ethnomethodology views social order as an achievement, and not as an objective pre-existing social fact that we uncover (Garfinkel, 1967).

Indexicality and *reflexivity* are two ideas that are very important in ethnomethodology and CA. Indexicality asserts that the meaning of words, texts, utterances, and silences depend on the context in which they are used (who is speaking, when they are speaking, and where they are speaking). In turn, reflexivity means that the talk is not a mere linguistic representation, a substitute of social world, but it is in itself a social reality, a social world.

In these two characteristics, ethnomethodology fits into the broader tradition of qualitative research, marked by *contextual understanding of action* and *constructivist ontological position*. In its detailed analysis of conversations recorded in naturally-occurring settings, CA also fits into another tradition of qualitative research, that for *naturalism* or analyzing naturally occurring situations. However, context in CA has much narrower meaning than in qualitative tradition in general: Context here means the specific situation immediately preceding the talk, and not the wider context of culture of the group, its values and beliefs, as in qualitative research. CA analysts want to avoid broader analysis of culture that is not grounded in participants' immediate concerns.

CA is based on several assumptions. It assumes the following:
1. *Talk is structured*: It follows patterns that are recognized and followed by participants. The structures of talk are revealed in the flow of conversation, pauses, and emphases.
2. *Talk is forged contextually:* It must be understood contextually and from the talk preceding it.
3. *Analysis should be grounded in data*: characteristics of talk and social order must be derived from data.

CA is therefore based on a very detailed analysis of speech, and attention to details is paramount here. CA researchers work with very detailed transcripts, where every pause, every intonation, every change of topic is noted. They look at recurring characteristics of conversations to examine their flow. Some of these recurrent features of conversations are turn-taking, adjacency pairs, and preferred vs non-preferred responses.

- **Turn-taking** shows that conversations are basically a collaborative exercise, usually following a pattern of turn-taking between two parties. CA analysts examine the signs that indicate that one's turn in a conversation is completed, since people implicitly share some codes or understanding about when one's utterance ends and the turn occurs for the other person to speak.
- **Adjacency pairs** indicate that talking activities are linked to each other, such as the question followed by an answer, invitation followed by a response and so on.
- Conversations are organized around preferred response (**preference organization**), for example, when acceptance of the invitation is a preferred response and declining an invitation is a non-preferred response. Participants of the conversation recognize that conversation is structured in this way, and feel the need to justify non-acceptance while the acceptance does not have to be justified. Therefore, the way in which the preferred responses are given is more straightforward, while the non-preferred responses require justifications, pauses, and awkward moments.

In summary, conversation analysis is a rather structural approach to analyzing action (conversation), and finds it illegitimate to invoke cultural factors when explaining conversations. Rather, the CA asks research questions that pertain to the structure and history of the talk itself, and explicitly refrains from discussion of cultural contexts or broader cultural factors that influence conversations.

Discourse Analysis

Discourse analysis (DA) is the study of how a view of the world or understanding of an object is produced through discourse. It includes elements of conversation analysis, but it is broader and more flexible. It covers communication beyond naturally-occurring talk, and looks at how linguistic categories shape people's understanding of the world. DA shows how the understanding of the studied problem emerged in historical and public interpretations, and how those in power shape a particular understanding of the problem. In sociology, it looks at how the relationships of power are reproduced in a discourse (in writings of Michel Foucault). The objects of DA are texts, conversations, and other sources of communication.

What are the philosophical assumptions behind the discourse analysis? Two main points describe epistemology and ontology of DA:
1. Discourse analysis tends toward an *anti-realist orientation*. That is, it assumes that there is no objective reality waiting to be found, and therefore no objective account of social world is possible.
2. Discourse analysis is also *constructivist*: It gives priority to the accounts produced by the actual participants, the members of social setting, and recognizes that many different interpretations of a situation are possible.

Given these characteristics, a DA is never a simple description of the situation. Rather, it is action oriented, and examines particular strategies people use to create different effects in communication. So a DA is targeted and usually asks the following questions (Potter, 2004):
- What is the discourse doing?
- How is it constructed to make this happen?
- What resources are available to perform this activity?

In this sense, discourse analysis performs a certain action, showing how the discourse is formed and how the parties involved in communication achieve their respective objectives. DA is similar to CA in that it looks at contextual understandings that emerge from situational specifics of the talk, but its main approach to analysis consists of "adopting a posture of sceptical reading" (Gill, 2000). Gill (2000) identified four key themes in discourse analysis:
1. *Discourse is a topic in itself*: it is not just a means of getting at a reality that lies behind it. Unlike qualitative researchers who think of the language of the interview as a way to reveal

the real thoughts and behaviour of people, the language of the discourse is the topic of inquiry in itself.

2. *Language is constructive*: discourse analysis is a way of constructing a particular view of social reality, with its own choices in presenting the problem.
3. Discourse is *a form of action*: language is a way of accomplishing acts, such as advancing an argument, attributing blame, or presenting oneself in a particular way.
4. Discourse *is rhetorically organized*: it presents one version of the world in the face of competing explanations.

Discourse analysis is constructed using several techniques:

- It uses *quantification rhetoric*, referring to numbers and stats to support or refute the argument
- It uses the technique of *variation in numbers* to highlight contrast (e.g., contrasting percentages with absolute numbers)
- It gives *attention to details*, emphasizing the supportive details used in the discourse (e.g., emphasizing that the three most curable cancers are also the rarest forms of cancer)
- DA gives *attention to rhetorical detail* to construct convincing arguments, and makes effort to discount a possible counter-argument.

Critical discourse analysis deals with exposing the political nature of the examined texts, considering the issues of power hierarchies, structural inequalities, and historical political struggles. It shows how talks and texts manifest and reproduce ideologies, power, and inequality in language. Its main preoccupation is how the existing power relationships are reproduced and reinforced in the discourse. Critical discourse analysis is also committed to social change and empowerment of those oppressed in the discourse.

Some critical discourse analysts relate discourses to social structures that underpin them and show how discourses work in these social structures (Reed, 2000). In this approach the discourse is not conceived purely as the use of language, but is viewed through its links to broader social influences, making this strand of DA closer to social science than to analysis of language. In other words, critical DA considers how both social factors and linguistic practices influence power relations and shape the discourse.

Discourse analysis is similar to conversation analysis in its analysis of talk, turns of conversation, and intersubjective meanings. However, DA departs from CA in that it looks on motives of discourse. Conversation analysts object that discerning the motives of actors in discourse analysis is pretty arbitrary and depends on analyst's choice. Discourse analysts retort that the CA analyzes only a small part of social reality by looking at conversations, because conversations are only tiny fragments of social life.

Advantages of Content Analysis

- In its quantitative form it is very transparent, and therefore easily replicated.
- CA allows for longitudinal analysis, since media coverage can be studied over time.
- It is an unobtrusive method in that newspaper articles and TV scripts are not usually written in anticipation of CA carried out on them, while an analysis of qualitative interview transcript reflects a higher obtrusiveness and possible reactivity of a respondent in a qualitative interview.
- CA is flexible: It can be used with several kinds of unstructured information, and can be applied on broad range of media output.
- CA can overcome social barriers to researcher access. For example, a lot of CA portrayal of the elites, businessmen or top officials can come from newspaper coverage and specialized publications (*Who Is Who* and the like).

Disadvantages of Content Analysis

- There are limitations due to the text analyzed (e.g., authenticity, credibility, representativeness): CA is only as good as the text it studies.
- Even a quantitative CA relies on interpretations, and can produce inconsistencies between different coders.
- There is a potential for invalid conjecture, especially in discussions of latent meanings as opposed to manifest meanings in qualitative CA. Qualitative analysis of text is also prone to higher degree of discretion in interpretation than quantitative counting of words.
- It is difficult to answer "Why?" questions using content analysis. CA shows how different meanings and interpretations emerge in a text, but it does not show why it is the case.
- The emphasis of some CA on measurement makes it look atheoretical in nature, since it focuses on what is measurable rather on what is theoretically significant or important.

Learning Objectives

In this chapter, you should learn to do the following:
- Understand that the main purpose of *content analysis* is an analysis of different forms of communication: texts, talk, TV programs, websites, and others.
- Differentiate between qualitative and quantitative content analysis
- Understand and be able to apply the coding principles in quantitative content analysis, such as the form and content of the coding manual and potential problems with designing the coding scheme in content analysis
- Discuss the main purpose and principles of *semiotics* as the science of studying signs in communication, and *hermeneutics* as the study of historical texts which takes into account the perspective of text author and considers the context in which the text was produced
- List the main purposes and principles of *conversation analysis* as an analysis of talk which looks at underlying structures in interaction that make social order possible
- Discuss the main goals and principles of *discourse analysis*, which is an analysis of various forms of communication that looks at ways in which language can create versions of reality
- Pinpoint the specific features of *critical discourse analysis* as a method that exposes the political nature of the examined texts, considering the issues of power hierarchies, structural inequalities, and historical political struggles
- Summarize the advantages and disadvantages of content analysis

Media Resources

Chandler, D. Semiotics for Beginners.
http://visual-memory.co.uk/daniel/Documents/S4B/sem02.html
- How might an object be both the sign and signifier at the same time?
- Can the signifier and signified be separated?
- How do objects and actions co-relate with each other in semiotics?

Know Your Audience: Content Analysis
http://www.audiencedialogue.net/kya16a.html
- How broad is a range in material that content analysis can be used for?

- What is the difference in the process for content analysis for TV versus newspaper articles?
- What sort of variation in conclusion might you expect to see from different researchers who are doing content analysis?

Critical Discourse Analysis: A Primer
http://www.kon.org/archives/forum/15-1/mcgregorcda.html
- How does discourse analysis reach beyond conversation analysis?
- How does discourse analysis compare to conversation analysis?
- Compared to semiotics is discourse analysis an easier or harder research process to engage?