CHAPTER 2

Conceptual Foundation

Content analysis has its own approach to analyzing data that stems largely from how the object of analysis, content, is conceived. This chapter introduces an epistemologically grounded definition of content analysis that guides the remainder of the book. It discusses its relationship to other definitions, other methods of social inquiry, and where and how it deviates from them. A conceptual framework is developed in terms of which the purposes and processes of content analysis may be understood. It includes the researcher, the knowledge he or she needs to bring to it, and the criteria by which a content analysis can be justified, and it makes transparent the function of various methods to be discussed in subsequent chapters. This chapter concludes with the suggestion that the ongoing changes in information technology make content analysis an increasingly attractive method for understanding social phenomena.

2.1 Definition

Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use.

As a *technique,* content analysis involves specialized procedures. It is learnable and divorceable from the personal authority of the researcher. As a research technique, content analysis provides new insights, increases a researcher's understanding of particular phenomena, or informs practical actions. Content analysis is a scientific tool.

Techniques are expected to be *reliable.* More specifically, research techniques should result in findings that are *replicable.* That is, researchers working at different points in time and perhaps under different circumstances should get the same results when applying the same technique to the same phenomena. Replicability is the most important form of reliability.

Scientific research must also yield *valid* results, in the sense that the research effort is open to careful scrutiny and the resulting claims can be upheld in the face

of independently available evidence. The methodological requirements of reliability and validity are not unique to but make particular demands on content analysis.

The reference to *text* in the above definition is not intended to restrict content analysis to written material. The phrase "or other meaningful matter" is included in parentheses to indicate that in content analysis works of art, images, maps, sounds, signs, symbols, and even numerical records may be included as data—that is, they may be considered as texts—provided they speak to someone about phenomena outside of what can be sensed or observed. The crucial distinction between text and what other research methods take as their starting point is that a text means something to someone, it is produced by someone to have meanings for someone else, and these meanings therefore must not be ignored and must not violate why the text exists in the first place. Text—the reading of text, the use of text within a social context, and the analysis of text—serves as a convenient metaphor in content analysis.

In the content analysis literature, scholars have provided essentially three kinds of definitions of this research method:

1. Definitions that take content to be *contained* in a text
2. Definitions that take content to be *a property of the source* of a text
3. Definitions that take content to *emerge in the process of a researcher analyzing a text* relative to a particular context

Each of these kinds of definitions leads to a particular way of conceptualizing content analysis and, consequently, of its analytical procedures.

Berelson's original definition of content analysis is an example of the first kind. Berelson (1952) defined content analysis as "a research technique for the objective, systematic and quantitative description of the manifest content of communication" (p. 18). His requirement that content analysis be "objective" and "systematic" is subsumed under the dual requirements of replicability and validity in our definition. For a process to be replicable, it must be governed by rules that are explicitly stated and applied equally to all units of analysis. Berelson argued for "systematicity" in order to combat the human tendency to read textual material selectively, in support of expectations rather than against them. Our requirement of validity goes further, demanding that the researcher's processes of sampling, reading, and analyzing messages ultimately satisfy external criteria. Replicability is measurable and validity is testable, but objectivity is neither.

Our definition of content analysis omits three of Berelson's further requirements. One is his insistence that content analysis be "quantitative." Although quantification is important in many scientific endeavors, qualitative methods have proven successful as well, particularly in political analyses of foreign propaganda, in psychotherapeutic assessments, in ethnographic research, in discourse analysis, and, oddly enough, in computer text analysis. The ability of computers to crunch words as well as numbers is well-known. When a computer program is used to analyze words, the algorithms that determine the program's operation must

embody some kind of theory of how humans read texts, rearticulate texts, or justify actions informed by the reading of texts. Reading is fundamentally a qualitative process, even when it results in numerical accounts. By including the attribute "manifest" in his definition, Berelson intended to ensure that the coding of content analysis data be reliable; this requirement literally excludes "reading between the lines," which is what experts do, often with remarkable intersubjective agreement (I will have more to say on this topic later in this chapter).

My chief objection to Berelson's definition, and numerous derivatives of that definition, is related to his phrase "description of the manifest content of communication." It implies that content is obviously *manifest* in messages, waiting to be separated from its form and described. Berelson felt no need to elaborate on the crucial concept of "content" in his definition because for him and his contemporaries, at the time of his writing, there seemed to be no doubt about the nature of content—it was believed to reside *inside* a text.

Berelson's operationalization of the attribute "manifest" is telling. If sources, receivers, and content analysts have different interpretations of the same message, which is quite natural, Berelson's definition restricts content to what is *common to all of these accounts,* what everyone can agree to. Gerbner (1985) starts from a similar assumption when he insists that mass-media messages carry the *imprint* of their industrial producers. For him, too, content is right there to be described for what it is. However, Gerbner goes beyond Berelson's notion by suggesting that the messages of the mass media are revealed in statistical accounts of their contents. Mass-media audiences, he suggests, are affected by certain statistical properties of mass-produced messages of which neither mass producers nor mass audiences are conscious. This privileges content analysts' accounts over the readings by audience members. Shapiro and Markoff's (1997) definition equates content analysis with scientific measurement as well, specifically, with "any systematic reduction . . . of text (or other symbols) to a standard set of statistically manipulable symbols representing the presence, the intensity, or the frequency of some characteristics relevant to social science" (p. 14). Its implicit representationalism is common in several definitions of content analysis. For example, Riffe, Lacy, and Fico (1998) start their textbook with the proposition that content is central to communication research but then assert that the purpose of content analysis is to describe "it" so as to make "it" amenable to correlations with other (noncontent) variables—as if content were a variable or thing inherent to mass-media messages. These examples demonstrate that the container metaphor for meaning still abounds in much of the communication research literature (Krippendorff, 1993). The use of this metaphor entails the belief that messages are containers of meaning, usually one meaning per message, and justifies calling any analysis of any conventionally meaningful matter a content analysis, regardless of whether it counts words or offers in-depth interpretations. Clearly, this is an insufficient way to define content analysis.

Definitions of the second kind distinguished above tie the content analysis of texts to *inferences about the states or properties of the sources of the analyzed texts* (Krippendorff, 1969a, p. 70; Osgood 1959, p. 35). These definitions tend to generalize the aims of psychological tests to reveal what is going on inside an individual: emotional states, cognitions, intentions, attitudes, and prejudices. Shapiro and Markoff (1997), among

others, have criticized such definitions as too limiting. Holsti (1969, p. 25) elaborates on this idea by committing content analysis to an encoding/decoding paradigm in which message sources are causally linked to recipients through processes of encoding, transmission through channels, and decoding by a receiver. Holsti wants the content analyst to describe the characteristics of messages in terms of "what," "how," and "to whom" in order to infer their antecedents in terms of "who" and "why" and their consequences in terms of "with what effects." These questions could be answered directly if sources and recipients were accessible to observation or able to inform the analyst honestly. When antecedents and consequences are not accessible to direct observation, the analyst must make inferences. I am sympathetic to Holsti's logic, but putting sources—senders and/or receivers—in charge of the validity of the inferences may not be the best way for the content analyst to capture all of the communicators' circumstances and intents. Moreover, describing message characteristics in terms of "what," "how," and "to whom" fails to acknowledge the analyst's own conceptual contributions to what constitutes the appropriate reading of the analyzed texts and the relevance of this reading to a given research question.

The analyst's conceptual contributions to the reading of a text are specifically recognized in a call for an *ethnographic content analysis* (Altheide, 1987); unfortunately, however, this call has not yielded specific methods of analysis. Proponents of ethnographic content analysis oppose the sequential nature of traditional content analysis, suggesting instead that analysts be flexible in taking into account new concepts that emerge during their involvement with texts. This call acknowledges the theory-driven nature of content analysis but also demands that the analytical process be closely linked to the communicators studied. Ethnographic content analysis is emic rather than etic in intent; that is, it attempts to rely on indigenous conceptions rather than on conceptions imposed by an analyst's theory. Although the preference for communicators' conceptions would appear to tie ethnographic content analysis to the second kind of definition noted above, by urging researchers to reflect on their involvement in the process, the approach acknowledges the possibility that researchers' theories can play a role in how analysis proceeds. The latter ties it more closely to the third kind of definition of content analysis, which we now explore.

Epistemological Elaborations 2.2

The definition of content analysis offered at the opening of this chapter is of the third kind. It focuses attention on the process of content analysis and does not ignore the contributions that analysts make to what counts as content. The key to the definition lies in the operations that define the nature of content analysis data. Most content analysts probably realize that the starting points of their analyses, *texts* (printed matter, recorded speech, visual communications, works of art, websites, and cultural artifacts), are quite unlike physical events in that they are meaningful to others, not just to the analysts. Recognizing meanings is the reason that researchers engage in content analysis rather than in some other kind of investigative method. A content analyst must acknowledge that all texts are produced and read by others and are expected to be significant to them, not just to

the analyst. Inasmuch as linguistically competent communicators are able to transcend the physical manifestations of their messages and respond instead to what those messages mean to them, content analysts cannot remain stuck in analyzing the physicality of text—its medium, characters, pixels, or shapes. Rather, they must look outside these characteristics to examine how individuals use various texts in their respective worlds. It would follow that the popular measurement model for conceptualizing content analysis, borrowed from mechanical engineering and widely used in the natural sciences and behavioral research, is misleading; it implies that there is nothing inherent in text that could be measured without interpretation by competent authors, readers, users, and—we need to include—culturally competent analysts. Below, I elaborate on six features of texts that are relevant to our definition of content analysis.

1. *Texts have no objective*—that is, *no reader-independent—qualities.* Seeing something as a text entails an invitation, if not a commitment, to reading it. Regarding something as a message implies that someone is trying to make sense of it. Accepting particular markers as data entails taking them as an unquestionable ground for subsequent conceptualizations. Thus texts, messages, and data arise in the process of *someone* engaging with them conceptually. A text does not exist without a reader, a message does not exist without an interpreter, and data do not exist without an observer. In a content analysis, it is methodologically trained researchers who, being familiar with their texts, design the analysis, instruct their coders to describe textual elements, and end up interpreting the results—always in the expectation of others' understanding. There is nothing inherent in a text; the meanings of a text are always brought to it by someone. Ordinary readers and content analysts merely read differently.

2. *Texts do not have single meanings* that could be "found," "identified," "described" for what they are, or correlated with states of their sources. Just as texts can be read from numerous perspectives, so signs can have several designations and data can be subjected to various analyses. One can count the characters, words, or sentences of a text. One can categorize its expressions, analyze its metaphors, describe the logical structure of its compositions, and ascertain its associations, connotations, denotations, and commands. One can also offer psychiatric, sociological, political, or poetic interpretations of that text. All of these accounts may be valid but different. Untrained analysts may be overwhelmed by these choices. Researchers who pursue content analysis according to the first of the above definitions are led to believe that a message has but one content, all other meanings being deviant, wrong, or subjective, and hence to be excluded. This naive belief is an entailment of the unreflecting use of the container metaphor. Perhaps the term *content analysis* was ill chosen for this reason. The possibility that any text may have multiple readings renders the frequently published claims by some researchers that they have analyzed *the* content of particular bodies of text untenable by our (third kind of) definition.

3. *The meanings invoked by texts need not be shared.* Although intersubjective agreement as to what an author meant to say or what a given text means to a community would simplify a content analysis tremendously, such consensus rarely exists in fact.

Demanding that analysts limit themselves to "common" or "shared ground" would restrict the empirical domain of content analysis to the most trivial or "manifest aspects of communications," on which Berelson's definition insists, or it would restrict the use of content analysis to a small community of message producers, recipients, and analysts who happen to see the world from a common perspective. If content analysts were not allowed to read texts in ways that are different from the ways other readers do, content analysis would be pointless. In fact, psychiatrists are expected to interpret the stories they hear from their patients in ways that differ from the patients' interpretations. Anthropologists' analyses of cultural artifacts need not conform to what informants say about those artifacts, and conversation analysts have good reasons to see verbal interactions in ways conversants might not. As Gerbner and his colleagues have shown through content analyses, mass-media audiences are not aware of the statistical trends in the qualities of popular heroes, the kinds of violence depicted, and the representations of minorities in television programming. Critical scholarship would be stifled if it could not go outside of what everyone accepts as true. Content analysis is in trouble only when expert interpretations fail to acknowledge the multiplicity of uses of texts by designated populations of readers or actors, particularly when content analysts fail to spell out the criteria for validating their results.

4. *Meanings (contents) speak to something other than the given texts,* even where convention suggests that messages "contain" them or texts "have" them. Probably the most distinctive feature of communications is that they inform their recipients, invoke feelings, or cause behavioral changes. Texts can inform their readers about events at distant locations, about objects that no longer exist, about ideas in other people's minds, about available actions—just as symbols represent things in their absence and stories walk their listeners through imagined worlds. Texts can also lead to responses of various kinds. All of these phenomena link the reading of present texts to something else, not necessarily present. Whether these other phenomena concern purely mental constructions, past or future experiences, or hidden causes, the analyst must be able to conceive of them and verbalize them. It follows that content analysts must look outside the physicality of texts—for example, to how people other than the analysts use these texts, what the texts tell them, the conceptions and actions the texts encourage. This requirement is a key to understanding the limitations inherent in computer text analysis. Computers can be programmed to manipulate character strings in amazingly complex ways, but their operations remain confined to the conceptions of their programmers. Without human intelligence and the human ability to read and draw inferences from texts, computer text analysis cannot point to anything outside of what it processes. Computers have no environment of their own making; they operate in the contexts of their users' worlds without understanding those contexts.

5. *Texts have meanings relative to particular contexts, discourses, or purposes.* Although diverse readings of a text are typical, the task of content analysts is far from hopeless. Messages always occur in particular situations, texts are read with particular intents, and data are informative relative to particular problems. Statisticians, linguists,

anthropologists, psychiatrists, and political analysts all have their own discipline-based reasons for interpreting given assertions differently. A therapist and a conversation analyst will view the same conversation differently. A speech on economics may be analyzed for its political implications, for how well it presents certain arguments, for what the speechwriter knows about economics, or for the emotions it arouses. We explain these differences by the *contexts* within which analysts choose to listen to that speech. Differences in interpretations do not preclude the possibility of agreements within particular contexts, however. In fact, once content analysts have chosen the context within which they intend to make sense of a given text, the diversity of interpretations may well be reduced to a manageable number, sometimes to one.

Every content analysis requires a context within which the available texts are examined. The analyst must, in effect, construct a world in which the texts make sense and can answer the analyst's research questions. A context renders perceptual data into readable texts and serves as the conceptual justification for reasonable interpretations, including for the results of content analysis. Often, analysts presuppose particular contexts based on their own disciplinary commitments, as in the above example about a speech on economics. Analysts working within particular disciplines, such as political science, rhetoric, economics, and psychology, hold particular theories concerning how texts are to be handled; that is, they are willing to accept only a certain context. Holsti's encoding/decoding paradigm, mentioned above, functions as a prominent analytical context in communication research, but it is by no means the only one. The contexts that psychiatrists tend to construct are very different from those that political scientists are likely to accept or within which literary scholars prefer to work. Once an analyst has chosen a context for a particular body of text and clearly understands that context, certain kinds of questions become answerable and others make no sense.

Just as the analytical contexts that content analysts must adopt may vary from one analysis to another, these contexts may also differ from the interpretive schemes that unaided listeners, viewers, or readers employ in reading their sensory data, the characters of their texts, and the messages they receive. The same body of texts can therefore yield very different findings when examined by different analysts and with reference to different groups of readers. For a content analysis to be replicable, the analysts must explicate the context that guides their inferences. Without such explicitness, anything would go.

6. *The nature of text demands that content analysts draw specific inferences from a body of texts to their chosen context*—from print to what that printed matter means to particular users, from how analysts regard a body of texts to how selected audiences are affected by those texts, from available data to unobserved phenomena. Texts, messages, and symbols never speak for themselves. They inform someone. Information allows a reader to select among alternatives. It narrows the range of interpretations otherwise available. For the content analyst, the systematic reading of a body of texts narrows the range of possible inferences concerning unobserved facts, intentions, mental states, effects, prejudices, planned actions, and antecedent or consequent conditions. Content analysts infer answers to particular research questions from their texts. Their inferences are merely more systematic, explicitly

informed, and verifiable—ideally—than what ordinary readers do with texts. Recognizing this apparent generality, our definition of content analysis makes the drawing of inferences the centerpiece of this research technique.

The element of "making inferences" is not entirely absent from other definitions of content analysis. For example, Stone, Dunphy, Smith, and Ogilvie (1966) define content analysis as "a research technique for making inferences by systematically and objectively identifying specified characteristics within a text" (p. 5). Although their inclusion of "within a text" here would suggest a commitment to "inherentist" conceptions of meaning, Stone et al. nevertheless recognize the inferential character of the processes of coding and categorizing textual material, in their case by computer. Their dictionary of fixed linguistic classifications of word meanings leads to semantically simplified representations of a text's conventional readings. Other authors have equated inferences with statistical generalizations (e.g., Roberts, 1997), which do not, however, move into the context of textual matter. As early as 1943, Janis (1943/1965) pointed to the need for researchers to validate the results of content analyses of mass communications by relating research findings to audience perceptions and to behavioral effects. Our definition requires that content analysts be able to validate their results as well, whether those results are used to predict something, to inform decisions, or to help conceptualize the realities of certain individuals or groups. But validation becomes an issue only where inferences are specific and thus have the potential for failing.

Regarding the drawing of inferences, Merten (1991) paraphrases the essential elements of my definition of content analysis (Krippendorff, 1980b) when he writes, "Content analysis is a method for inquiring into social reality, which consists of inferring features of a nonmanifest context from features of a manifest text" (p. 15; my translation). All theories of reading (hermeneutics) and theories of symbolic forms (semiotics), including theories of message meanings (communication/conversation theory), can be operationalized as processes of moving from texts to the contexts of the texts' use. I would also suggest that a context is always constructed by someone, here by content analysts, no matter how hard they may try to objectify it. This is true even for ethnographers who believe that they can delegate the definition of the context to their informants' world conceptions. It is the ethnographers who are held responsible for what they end up reporting. One cannot deny content analysts' interest and conceptual participation in what their analysis reveals. Whether the analysts' context coincides with the many worlds of others is a difficult question to answer. Whether the analysts' world makes sense to their scientific peers depends on how compellingly the analysts present that world.

Examples 2.3

In this section, I offer some examples to illustrate how our definition of content analysis applies to practical situations.

Example 1. Consider the situation of wartime analysts of enemy broadcasts who want to gauge, among other phenomena, the popular support that enemy elites

enjoy in their country. In peacetime, researchers could obtain such information directly, through public opinion surveys, for example, or by on-site observations. In wartime, however, information of this nature is difficult to get, if not deliberately concealed, and analysts are forced to use indirect means of obtaining it. The inability to use direct observation is an invitation to apply content analysis. Here, analysts are typically not interested in the literal meanings of enemy broadcasts, in the rhetorical devices political leaders use, or in judging whether individual citizens are being deliberately misled. In fact, wartime propaganda analysts have good reasons to overlook manifest contents and ignore their truths. To infer from enemy domestic broadcasts the extent of popular support for elite policies, the analysts must understand that the broadcasts are part of a complex communication network in which the mass-media system and political system interact with a population to make news acceptable. The propaganda analysts have to know something about the actors involved in the governing elite and in the military, about the media these actors have access to, and about other institutions that have a stake in current affairs. They must also have some knowledge of the political-economic processes that keep a country together and how the public tends to respond to mass-mediated messages. The picture they construct of what they are dealing with amounts to the context of their analysis. It connects the intercepted broadcasts to the phenomena of interest, whether they concern popular support of the governing elite's policies, planned military actions, or evidence of war weariness.

Example 2. Historians are never mere collectors of documents. They offer reconstructions of past events that they deem consistent with current readings of all available documentary evidence. Historians are far removed from the worlds they wish to articulate. They cannot interview Julius Caesar, ask Homer about his sources for the *Iliad,* participate in the experiences of African slaves entering colonial America, or listen to conversations between Pablo Picasso and Henri Matisse. Historical figures reside in our readings of available documents, not in facts. And although some have left their writings to us, it is unlikely that they anticipated contemporary historians' readings. Past happenings become comprehensible to us only by inferences from documents that have survived to the present (Dibble, 1963). Historians who infer past events from available texts are, by our definition, involved in content analysis. It is not surprising, therefore, that historians are keenly aware of the need to place the documents they analyze within the context of other relevant documents. Without the appropriate context, a document means very little; a document placed in the wrong context acquires incorrect meanings, or at least meanings that may not make much sense. Historiographical methods organize available documents into webs of inferential relationships that may ultimately answer a historian's questions.

Example 3. Psychological researchers have a long tradition of developing theories whose generalizability is established by repeated experiments. The subjects of psychological research must be present, however, making it difficult for researchers to study developmental issues and individuals who are available only through their writings. Expanding psychological research methods, Allport (1942) added personal

documents, witness accounts, and letters to the repertoire of data amenable to psychological inquiries. The research he proposed amounts to content analysis by our definition: There are texts in the form of personal documents, diaries, letters, and recorded speeches, and researchers construct the contexts for analyzing these texts with the help of available theories concerning the correlations between what people say and a variety of psychological variables (e.g., cognitive processes, attitudes, emotional arousal, personality traits, worldviews, or psychopathologies). Different schools of psychology direct their researchers to different questions, but they all are interested in inferring psychological variables of authors from the texts they left behind. In the course of analyzing personal documents, psychologically oriented content analysts have developed a variety of inferential techniques (e.g., type/token ratios of key concepts, the discomfort/relief quotient, graphological interpretations, readability yardsticks, Thematic Apperception Tests, and personal structure analysis). In individual psychology, content analysis has become an established method of inquiry since Allport's (1965) pioneering work.

Example 4. For good reasons, interview and focus group data are frequently subjected to content analysis. Structured interviews generate predefined question-answer pairs, and the researcher then analyzes their distribution. The researcher's conceptions are imposed on the interviewees, who cannot express the reasons for their choices among predefined answers and whose individual conceptions are ignored. In open-ended interviews and focus groups, in contrast, participants are allowed to speak freely and in their own terms. To explore the conceptions that are manifest in such conversations, researchers need to perform what amounts to content analysis on the transcripts of these conversations. In a breast cancer study, for example, patients were asked about their lives after they had received treatment (Samarel et al., 1998). The answers were naturally freewheeling, as expected, enabling the researchers to adapt their theory of "coping" to the transcripts at hand. The researchers' reformulated theory then provided the context for a subsequent content analysis. Armed with questions derived from the researchers' theory, coders looked for and identified answers within the transcripts, and by tabulating these, the researchers provided frequencies and statistical accounts that the funders of the research required. In this study, the qualitative inferences were made during the process of coding, not based on the resulting frequencies, which merely summarized these inferences.

Example 5. Mass communication is the archetypal domain of content analysis. Communication researchers tend to be interested in communicator conceptions, media biases and effects, institutional constraints, implications of new technologies, audience perceptions, public opinion, and how certain values, prejudices, cultural distinctions, and reality constructions are distributed in society—relying on mass-media messages as their causes or expressions. Typically, mass-media material calls for more reading than any single person can handle. Its analysis thus requires a framework, a theory, a vocabulary, and an analytical focus in terms of which the researcher can construct a suitable context for analysis and collaborate with other researchers on the same project. Different contexts answer different research questions, of course.

A stereotypical aim of mass-media content analysis is to describe how a controversial issue is "depicted" in a chosen genre. Efforts to describe how something is "covered" by, "portrayed" in, or "represented" in the media invoke a picture theory of content. This approach to content analysis and decontextualizes the analyzed text and thus reverts to the first kind of definition of content analysis distinguished above. It conceals the researchers' interest in the analysis and hides their inferences behind the naive belief that they are able to describe meanings objectively while rendering the results immune to invalidating evidence. Consider common findings of political biases, racial prejudices, and the silencing of minorities on television as such issues. Although counts of evident incidences of such phenomena can give the impression of objectivity, they make sense only in the context of accepting certain social norms, such as the value of giving equal voice to both sides of a controversy, neutrality of reporting, or affirmative representations. Implying such norms hides the context that analysts need to specify. Unless analysts spell out whose norms are applied, whose attitudes are being inferred, who is exposed to which mass media, and, most important, where the supposed phenomena could be observed, their findings cannot be validated. Berelson and Lazarsfeld (1948, p. 6) noted long ago that there is no point in counting unless the frequencies lead to inferences about the conditions surrounding what is counted. For example, counting the numbers of mentions of *Microsoft* or *AIDS* or the term *road rage* over time in, say, the *New York Times* would be totally meaningless if the observed frequencies could not be related to something else, such as political, cultural, or economic trends. That something else is the context that lends significance to quantitative findings.

Example 6. Content analysis has many commercial uses. For example, word-association databases (which collect huge numbers of pairs of words that consumers associate in their minds, as determined through word-association experiments) can serve as the context within which advertising researchers can infer chains of associations for new products, services, or brand names. In another, very different application, Michael Eleey and I studied how publicity generated by the Public Broadcasting Service about its programming ended up in newspaper articles (Krippendorff & Eleey, 1986). The purpose of the study was to enable PBS analysts to infer how the Public Broadcasting Service is perceived by newspaper editors in different regions of the United States and to assess the effectiveness of PBS's publicity efforts. Here the context was very simple. It included what we knew about newspaper editors' access to wire services and press releases, their newspapers' coverage of PBS programming, and certain theories and assumptions about the difference between the two, which led us to infer the (controllable) persuasive force of PBS publicity and the (uncontrollable) attitudes and competencies of the journalists, further differentiated by region and size of the newspaper.

The foregoing suggests that purely descriptive intents, manifest in claims to have analyzed "*the content* of a newspaper," to have quantified "*the* media *coverage* of an event," or to have "*found* how an ethnic group is *depicted,*" fail to make explicit the very contexts within which researchers choose to analyze their texts. Content analysts have to know the conditions under which they obtain their texts, but, more important, they also have to be explicit about *whose* readings they are speaking

about, *which processes or norms* they are applying to come to their conclusions, and *what the world looks like* in which their analyses, their own readings, and their readings of others' readings make sense to other content analysts. Explicitly identifying the contexts for their analytical efforts is also a way of inviting other analysts to bring validating evidence to bear on the inferences published and thus advance content analysis as a research technique. The framework presented in the next section is intended to help content analysts to conceptualize the analytical process so that their results are arguably acceptable.

Framework 2.4

The definition of content analysis offered at the opening of this chapter and illustrated with the above examples emphasizes the drawing of inferences of a certain kind. It also assigns content analysts a particular role vis-à-vis their objects of inquiry. Following from the above and previous work (Krippendorff, 1969b, pp. 7–13; 1980b), I offer a conceptual framework for content analysis within which their role becomes clear and specialized methods can be located. This framework is intended to serve three purposes: Its *prescriptive* purpose is to guide the conceptualization and design of practical content analytic research; its *analytical* purpose is to facilitate the critical examination and comparison of the published content analyses; and its *methodological* purpose is to point to performance criteria and precautionary standards that researchers can apply in evaluating ongoing content analyses. Thus the use of the framework will lead to long-term systematic improvements of the method.

The framework, which is depicted in Figure 2.1 and will be revisited in Chapter 4's discussion of the logic of content analysis designs, is simple and general, employing only a few conceptual components:

- A body of text, the data that a content analyst has available to begin an analytical effort
- A research question that the analyst seeks to answer by examining the body of text
- A context of the analyst's choice within which to make sense of the body of text
- An analytical construct that operationalizes what the analyst knows about the context of the body of text
- Inferences that are intended to answer the research question, which constitute the basic accomplishment of the content analysis
- Validating evidence, which is the ultimate justification of the content analysis

2.4.1 Texts

Data are the starting point of any empirical research. Data are *taken as givens*—that is, the researcher is not in doubt as to what they are. In surveys, focus groups, and psychological experiments, researchers attempt to control the generation of their data, thereby assuring that they know what the data mean, largely, if not exclusively,

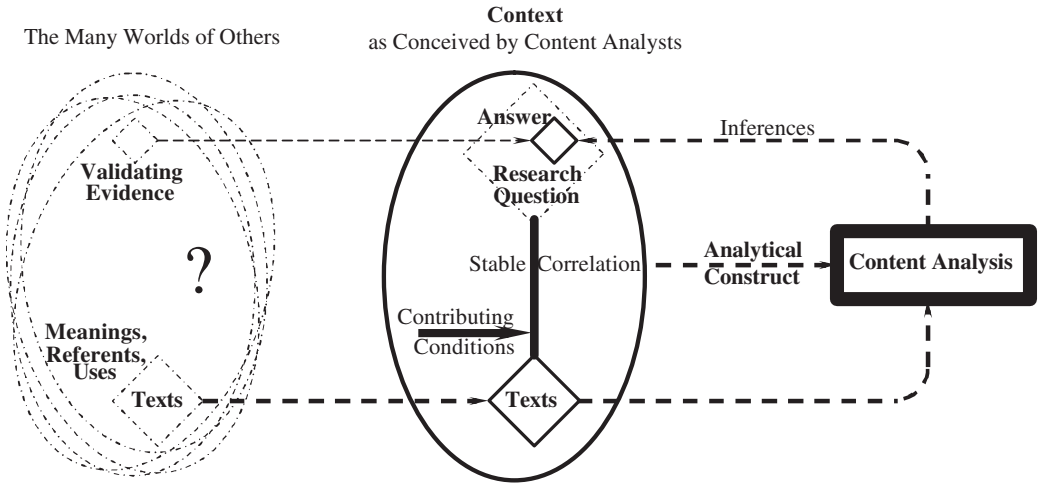


Figure 2.1 A Framework for Content Analysis

in the researchers' terms. Most content analyses start with data that are not intended to be analyzed to answer specific research questions. They are *texts* in the sense that they are meant to be *read, interpreted, and understood by people other than the analysts.* Readers may decompose what they read into meaningful units, recognize compelling structures, rearticulate their understandings sequentially or holistically, and act on them sensibly. When we are capable of this kind of rearticulation, we attribute textuality to what we see as writing, pictorial images, gestures, Web pages, videos, musical compositions, even behavioral sequences. Text results from reading and rearticulation.

One could speak of symbolic qualities instead of text, but it is preferable not to assume such qualities to exist without reference to who regards them as such. An analyst's reading—the units, syntax, and narrative structures that constitute the texts for the analyst—naturally differs from the readings that initiate the interpretations of ordinary readers, including the texts' authors. It follows that an analyst's reading must never be taken as the only legitimate one, nor should content analysts assume the sole power to determine the form of the texts they analyze. They would then be examining only themselves. We presume that all authors write in the expectation of being understood by self and by others, and it is the implication of others that renders a text socially significant. Although content analysts are not bound to analyze their data with reference to the conceptions or intended audiences of their texts' authors, they must at least consider that texts may have been intended for someone like them. We know that interviewees answer questions differently when they know how the research findings could affect them, and so we need to read interview results in the context of possible self-interests. We know that when politicians speak, they anticipate being scrutinized by the public, and so we cannot take their speeches at face value, as natural objects. Content analysts have to acknowledge that the textuality they rely on is not the only one that counts.

Content analysts' best guarantee against the contamination of texts by the stakes their sources have in how their texts are analyzed is to focus on textual features of which their sources are unconscious, or to apply categories that the sources of their texts use habitually or are unable to control. This is most obviously possible when the sources of texts are of the past (historical), when they are unaware of how their texts are being analyzed, or when communication to the analysts is one-way, without feedback. However, given that the results of most content analyses are published, and that the categories that analysts use have the potential of becoming known to the text sources as well, content analysts are justified in applying unconventional categories, that is, in looking at textuality in ways others may not. As Figure 2.1 illustrates, texts occur in the analyst's world but acknowledge their origins in the worlds of others.

2.4.2 Research Questions

Research questions are the targets of the analyst's inferences from available texts. Generally, such questions delineate several possible and initially uncertain answers. In this respect, a research question is analogous to a set of hypotheses. However, in contrast to scientific hypotheses, which are pitted against direct observational evidence, the research questions of content analysis must be answered through inferences drawn from texts. The difference between testing scientific hypotheses and selecting an answer to a research question is crucial. Whereas observations are registered or measured for what they are and hypotheses about observational phenomena amount to generalizations from observations, texts inform an analyst about extratextual phenomena, about meanings, consequences, or particular uses. Thus, whereas scientific hypotheses are accepted on account of a preponderance of evidence in favor of one at the expense of other hypotheses, an ideally large number of observations that support one and rule out others, inferences from texts (although large numbers may play a role here as well) pertain to phenomena that are not observed during a content analysis, phenomena that are outside the texts and thus retain their hypothetical character until confirmed by validating incidences.

There are two reasons for content analysts to start with research questions, ideally in advance of undertaking any inquiries: *efficiency* and *empirical grounding*. One can surely explore the meanings that come to mind while reading a text, following the threads of the inferences to wherever they may lead, or engaging in so-called fishing expeditions. Hermeneutical, interpretive, and ethnographic approaches to reading cherish such open-endedness. However, when research is motivated by specific questions, analysts can proceed more expeditiously from sampling relevant texts to answering given questions. Content analysts who start with a research question read texts for a purpose, not for what an author may lead them to think or what they say in the abstract.

The pursuit of answers to research questions also grounds content analysis empirically. All answers to research questions entail truth claims that could be supported, if not by direct observation then at least by plausible argumentation from

related observations. Our framework suggests that content analysis compensates for analysts' inability to observe phenomena in which they are interested, whether these phenomena pertain to the characteristics of writers or readers, to happenings hidden behind intentional information barriers, or to events in a distant past or future.

Formulating research questions so that the answers could be validated in principle protects content analysts from getting lost in mere abstractions or self-serving categorizations. For example, the question of how frequently a particular word occurs in a text can be answered by counting. Counting is what analysts do. Counts cannot be validated by independent evidence; to assure that counts are correct, analysts must repeat them, perhaps employing different persons as counters. The same is true for questions concerning whether one can categorize, measure, or analyze something. Their answers lie in a researcher's ability to execute these processes reliably. These questions cannot be answered by research. Questions concerning the statistical generalizability of textual attributes or "contents" (in the sense of the first kind of definition of content analysis discussed above) from a sample to a population from which this sample was drawn are not suitable content analysis research questions either, but for a different reason. Although their answers do rely on empirical evidence, without abductive inferences to phenomena outside the texts being analyzed, generalizations are inductive and cannot answer content analysis research questions. Thus, in content analysis, research questions have the following characteristics:

- They are believed to be answerable (abductively inferable) by examinations of a body of texts. (In Figure 2.1, this is indicated by the bold dashed arrows.)
- They delineate a set of possible (hypothetical) answers among which analysts select. (In Figure 2.1, an answer is indicated by the unlabeled diamond.)
- They concern currently inaccessible phenomena.
- They allow for (in)validation—at least in principle—by acknowledging another way to observe or substantiate the occurrence of the inferred phenomena. (In Figure 2.1, this is indicated by the thin dashed arrow from the worlds of others to the answer to the research question.)

2.4.3 Context

I have argued above that texts acquire significance (meanings, contents, symbolic qualities, and interpretations) in the contexts of their use. Although data enter a content analysis from outside, they become texts to the analyst within the context that the analyst has chosen to read them—that is, from within the analysis. A context is always someone's construction, the conceptual environment of a text, the situation in which it plays a role. In a content analysis, the context explains what the analyst does with the texts; it could be considered the analyst's best hypothesis for how the texts came to be, what they mean, what they can tell or do. In the course of a content analysis, the context embraces all the knowledge that the analyst applies to given texts, whether in the form of scientific theories, plausibly argued propositions, empirical evidence, grounded intuitions, or knowledge of reading habits.

The context specifies the world in which texts can be related to the analyst's research questions. This world is always one of many. Political analysts construct worlds that differ from those of politicians, often embracing additional perspectives, but those worlds also differ from the worlds of psychologists, journalists, historians, psychotherapists, scholars of literature, and—naturally—communication researchers, who pursue their own research agenda and approach texts with their own questions, concepts, models, and analytical tools. Scholars in different disciplines tend to place the same texts in different contexts, but rarely without acknowledging that there are other readings, other contexts, other worlds, within which given texts function as well—authors, audiences, users, and beneficiaries, for example. In Figure 2.1, these worlds are shown in the ovals embracing texts and their multiple meanings.

Knowledge of the context for content analyzing given texts can be separated into two kinds:

- The network of stable correlations, which are believed to connect available texts to the possible answers to given research questions, whether these correlations are established empirically, derived from applicable theory, or merely assumed for the purpose of an analysis
- Contributing conditions, which consist of all the factors that are known to affect that network of stable correlations in foreseeable ways

In Figure 2.1, these relationships are shown by a bold line and a bold arrow.

To use an example that is far from simple: In an ordinary conversation, what is observed and heard as being said at any one moment (the data) is understandable only in the context of what has been said before, by whom and to whom, the responses it elicited from the participants, and how it directed the conversation. This is an observer's account of a conversation, from outside of it. To participants, their version of what is going on (the contexts that include the other participants) is not necessarily shared. In fact, there would be no point in conversing if all participants saw their worlds, thought, and spoke alike. A conversation analyst contextualizes the transcript of a conversation (the text) in yet another way, by constructing a world (the analyst's context) within which the participants appear to "speak" in the analytical terms that the conversation analyst is familiar with and brings to the analyzed transcript. Whether a conversation analyst wants to infer the intentions of the participants to initiate certain moves (turn taking, for example) or how addressees will respond to a string of "he said–she said" (the evolution of a topic), the analyst draws on knowledge of the empirical relationship between these speech acts (the correlations that connect one to another) and the strengths (perlocutionary forces) of particular utterances, the network of connections that leads, hopefully, from texts to answers to the research question.

A conversation is not a mechanical system. Participants alter the rules of their engagement as it unfolds. This leaves outside observers uncertain as to what the participants mean, how they understand what is going on, and which rules govern the conversation at any one moment. Because conversation analysts tend not to participate in the conversations they analyze, and therefore have no way of asking

the interlocutors how they see their situation, the analysts have to acknowledge other determining variables (the contributing conditions) and find ways to ascertain how they affect the correlations relied upon to lead to the intended inferences.

Inasmuch as a context stands in place of what is momentarily inaccessible to direct observation, there is no limit to the number of contexts that may be applicable in a given analysis. Unless told, readers of the conclusions of a content analysis may not know the context that the analyst was using and may come to seriously misleading interpretations. In view of this possibility, *content analysts need to make their chosen contexts explicit, so that the results of their analyses will be clear to their scientific peers and to the beneficiaries of the research results.* Without explication of the context, the steps that a content analyst takes may not be comprehensible to careful readers, and the results to which they lead may not be validatable by other means.

2.4.4 Analytical Constructs

Analytical constructs operationalize what the content analyst knows about the context, specifically the network of correlations that are assumed to explain how available texts are connected to the possible answers to the analyst's questions and the conditions under which these correlations could change. Analytical constructs represent this network in computable forms. Extracted from the known or assumed context and entered into the research process, analytical constructs *ensure that an analysis of given texts models the texts' context of use,* which means that the analysis does not proceed in violation of what is known of the conditions surrounding the texts. Procedurally, analytical constructs take the form of more or less complex "if-then" statements, much like those used in computer programs. These "if-then" statements amount to *rules of inference* that guide the analyst, in steps, from the texts to the answers to the research questions. They also render knowledge of the context *portable to other content analyses* of similar contexts and make it possible for students and critics to examine the procedures that a content analyst has been using. In this respect, analytical constructs function much like testable mini-theories of a context, with the provision that they are computable on the coded features of available texts.

For example, a computer-aided content analysis might employ a dictionary of tags that mimics how competent speakers of a language categorize words into classes with similar meanings. Such a dictionary assumes linguistic stability, which may not be warranted, but it at least models a standard competence of language use. Another approach that an analyst might take is to adopt a computational theory of a context—a neuronal network model, for instance—that promises to explain how people form categories from words that occur in proximity to each other. Of course, labeling an analytical construct a "model" does not guarantee that it accurately represents the network of relationships that are relevant to readers and writers. More often, content analysts draw on empirically obtained correlations between observed and currently unobserved variables. Correlations measure the extent of a linear relationship between variables—for example, between the rate of

recorded speech disturbances and anxiety—which, if sufficiently general, could in turn be applied to individual cases, here yielding a prediction of a speaker's anxiety. However, as linguistic variables are rarely describable in intervals and linear regression equations tend to hold only under restricted conditions, the use of such constructs typically requires that the analyst have additional information about the conditions under which the construct is predictive of that behavior. Similarly, knowing that public agendas are influenced by the mass-media coverage of pertinent events may give a content analyst the idea of an analytical construct for analyzing media coverage in place of public opinion surveys. Such research, which has been done, requires a fairly detailed operationalization of the conditions under which verbal or pictorial elements influence particular public conversations.

Analytical constructs need not be perfect, of course, but unfortunately, many text analysts employ computational procedures that have no obvious relationship to any context in which given texts would arguably make sense. Counting units of text or applying sophisticated statistical techniques will always yield something, but this does not guarantee that the results will refer to anything. Content analysts must make sure that their analytical constructions model the contexts they have chosen. The purpose of all analytical constructs is *to ensure that texts are processed in reference to what is known about their use.*

2.4.5 Inferences

The inferential nature of content analysis should by now be obvious. Content analytic inferences may be hidden in the human process of coding. They may be built into analytical procedures, such as the dictionaries in computer-aided text analyses or well-established indices. Sometimes, especially after complex statistical procedures have been applied, inferences appear in the analyst's interpretations of the statistical findings. Figure 2.1 depicts the path that an inference takes with bold and broken lines, with the inference motivated or explained by an analytical construct that enters the analysis as a representation of the chosen context.

Because the word *inference* has several meanings, it is important to distinguish the meaning that is relevant to this discussion from others that are perhaps more familiar to readers. In logic, at least three types of inferences are distinguished:

- Deductive inferences are implied in their premises. For example, if all humans speak a language, then John, being human, must speak one as well. Deductive inferences are logically conclusive. They proceed from generalizations to particulars.
- Inductive inferences are generalizations to similar kinds. For example, I might infer from the fact that all of my neighbors speak English that all humans do. This inference is not logically conclusive, but it has a certain probability of being correct. Statistical generalizations from smaller samples to larger populations (typical of social research) and the idea of measuring the statistical significance of scientific hypotheses involve inferences of this kind. They proceed from particulars to generalizations.

- Abductive inferences proceed across logically distinct domains, from particulars of one kind to particulars of another kind. (These are the kinds of inferences of interest to content analysis, where they proceed from texts to the answers to the analyst's questions.) Consider linguistic competence and age. Logically, neither implies the other. However, if one has practical experience with infants' language acquisition, one might be able to infer children's ages from the sounds they make or from the vocabulary they use. Of course, one can make such inferences only with a certain probability, but the probability may be strengthened if one is able to take other variables (contributing conditions) into account.

Deductive and inductive inferences are not central to content analysis. The following examples of inferences employed in content analysis are all abductive in nature:

- One might date a document from the vocabulary used within it.
- One might infer the religious affiliations of political leaders from the metaphors used in their speeches.
- One might infer the readability of an essay from a measure of the complexity of its composition.
- One might infer whether someone is lying from his or her nonverbal (facial) behavior.
- One might infer the problems of a city from the concerns expressed in letters written to the city's mayor's office.
- One might infer the prevailing conceptualizations of writers and readers from the proximities of words in frequently used texts.
- One might infer editorial biases from a comparison of the editorial pages of different newspapers.
- One might infer a writer's psychopathology from the images used in her prose.
- One might infer the identity of the author of an unsigned document from the document's statistical similarities to texts whose authors are known.
- One might infer the political affiliations of citizens from the TV shows they choose to watch.
- One might infer an individual's propensity to engage in a hate crime from the ethnic categories he uses in ordinary speech.
- One might infer the likelihood of war from the coverage of international affairs in the elite newspapers of neighboring countries.

According to Eco (1994):

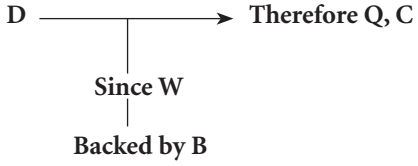
> The logic of interpretation is the Peircean logic of abduction. To explain a conjecture means to figure out a law that can explain a Result. The "secret code" of a text is such a Law. . . . in the natural sciences the conjecture has to try only the law, since the Result is under the eyes of everybody, while in textual interpretation only the discovery of a "good" Law makes the Result acceptable. (p. 59)

For Josephson and Josephson (1994, p. 5), abduction starts with a body of data (facts, observations, givens)—our text. A hypothesis—our analytical construct—if

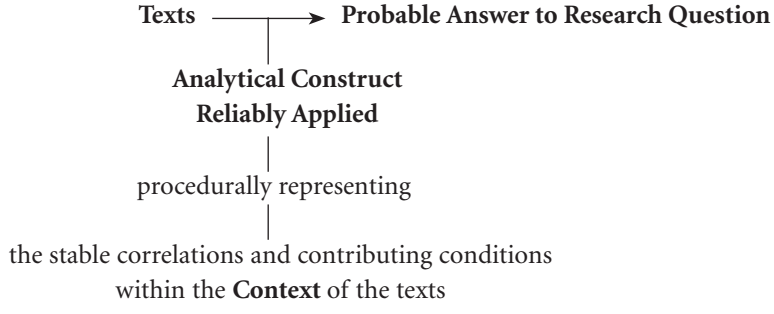
true, would explain these data. No other hypothesis can explain the data as well as the chosen one does. Therefore, the hypothesis is probably true and can be used to deduce other entailments—that is, answer our research questions.

Abductive inference is Sherlock Holmes's logic of reasoning as well (Bonfantini & Proni, 1988; Truzzi, 1988). Holmes's creator, Sir Arthur Conan Doyle, always lets him find empirical connections and apply bits of common knowledge in the context of established facts that he is then able to weave ingeniously into an inferential network containing the initially unrecognizable chain of logical steps from known facts to the perpetrator of an unobserved crime. Content analysts are in a similar position of having to draw inferences about phenomena that are not directly observable, and they are often equally resourceful in using a mixture of statistical knowledge, theory, experience, and intuition to answer their research questions from available texts.

In this respect, the whole enterprise of content analysis may well be regarded as an argument in support of an analyst's abductive claims. In Toulmin's (1958) theory of argumentation, which applies not just to abductions, the move from data (D) to conclusions or claims (C) must be justified by a suitable warrant (W). In his example, learning that "X is a Swede," the inference that "X most likely is a Protestant" is warranted by the knowledge that "most Swedes are Protestants." Because this inference is not without exceptions, it includes a qualification (Q) of the conclusion (C) (i.e., "most likely"). The warrant provides the logical bridge between the data and the conclusion. Toulmin also introduces another element: the ground on which the warrant may be justified, or the backing (B). In Figure 2.1 we may recognize the diagram that Toulmin (p. 104) uses to show the relationships among the above-mentioned parts of arguments:



In moving from texts to the answer to a research question, as illustrated in Figure 2.1, it is the assumptive analytical construct plus the assurance that the analysis has been performed reliably that warrants that inference, which in turn is backed by the analyst's knowledge of the context in which the texts occur or are interpreted:



2.4.6 Validating Evidence

Any content analysis should be validatable in principle. Because the raison d'être of content analysis is the absence of direct observational evidence, validation may be difficult or infeasible, if not impossible, in practice. It is infeasible when a content analysis is to be acted upon in the absence of direct observational evidence—for example, in wartime analysis of planned military activities from domestic propaganda or in assessments of whether a politician is lying. It is impossible when research questions concern past or future happenings, such as inferences from surviving documents to historical facts, inferences from the works of deceased authors to their intentions, or inferences from psychological tests to an individual's aptitude for a particular job. The point of requiring that content analyses be "validatable in principle" is to prevent analysts from pursuing research questions that allow no empirical validation or that yield results with no backing except by the authority of the researcher. For example, a conclusion from an analysis of television fiction that hedonism is on the rise in the United States means nothing unless those who claim such findings can show that this conclusion is not merely their abstraction from fictional programming, but also has some independently observable reality—that is, unless they can show that a rise in hedonism is manifest in something other than television fiction.

Ex post facto validation of content analysis is not merely a matter of curiosity. It can increase confidence in the results of future content analyses of similar texts and in similar contexts, but only if the categories of analysis and the analytical constructs are used repeatedly, so that successes and failures can be weighted against each other and used to advance the technique in the long run. Much too often, researchers design content analysis studies ad hoc and conduct them without any thought of validation; such research contributes little to the literature on content analysis.

A good example of ex post facto validation is George's (1959a) effort (mentioned in Chapter 1) to examine documents captured after World War II to see whether they matched what the Federal Communications Commission propaganda analysts had inferred during the war and to evaluate the FCC researchers' techniques for use by future analysts. In 1943, Janis (1943/1965) proposed an indirect method of validation, suggesting that the results of mass-media content analyses should at least correlate with audience verbal reports or observed behaviors (e.g., public opinion polls, voting, consumption, or aggression). Thus Gerbner and his colleagues sought to correlate the amount of violence seen on television with survey data on audience members' perceptions of how violent their world "really" is (see, e.g., Gerbner, Gross, Signorielli, Morgan, & Jackson-Beeck, 1979; Gerbner, Gross, Morgan, & Signorielli, 1994).

As noted above, our framework demands merely that a content analysis be validatable *in principle.* For example, if a content analyst infers what a particular group of voters learned from TV campaign ads about a candidate for political office, then, potentially, a survey of those exposed to the ads could validate or invalidate this inference. However, if a content analyst insists that such TV campaign ads have certain contents, there is no way for others to validate this "finding." Repeating this

particular content analysis would merely indicate the degree to which the original analysis was reliable. Similarly, finding that a certain word occurs with a certain frequency does not constitute an abductive inference. Recounting cannot validate what a frequency is inferred to mean.

Contrasts and Comparisons 2.5

Every research technique has its powers and its limitations, and content analysis is no exception. A researcher can misapply a technique or use a technique that is ill suited for a particular purpose, in ignorance of better ones. In this section, I contrast content analysis with other techniques used in social research, paying special attention to the four distinguishing features of content analysis.

Content analysis is an unobtrusive technique. As Heisenberg's uncertainty principle tells us, acts of measurement interfere with the phenomena being assessed and create contaminated observations; the deeper the observer probes, the greater the severity of the contamination. For the social sciences, Webb, Campbell, Schwartz, and Sechrest (1966) have enumerated several ways in which subjects react to being involved in scientific inquiries and how these can introduce errors into the data that are analyzed:

- Through the subjects' awareness of being observed or tested
- Through the artificiality of the task or the subjects' lack of experience with the task
- Through the expectations that subjects bring to the role of interviewee or respondent
- Through the influence of the measurement process on the subjects
- Through stereotypes held by subjects and the subjects' preferences for casting certain responses
- Through experimenter/interviewer interaction effects on the subjects

Controlled experiments, interviews, focus groups, surveys, and projective tests are especially vulnerable to such errors. By contrast, content analyses, computer simulations, research using already available statistics, and interpretive research (in cultural studies, for example) are nonreactive or unobtrusive. Researchers using ethnographic methods subscribe to the unobtrusive ideal as well, but while conducting fieldwork even the most careful ethnographers cannot escape influencing their informers.

Social researchers may want to avoid reactive situations for two primary reasons. The first is that *undue influence on the situation that gives rise to the data may distort the data, jeopardizing the validity of the research.* For this reason, ethnomethodologists prefer to obtain data in natural settings, psychiatrists avoid asking their patients questions that might induce false memories, and economists investigate mathematical models rather than experiment with the real economy. The second reason is that *researchers need to conceal their interest in the data for fear of being*

manipulated by their sources. Instrumental assertions are difficult to analyze (Mahl, 1959). Had Goebbels, the Nazi-era minister of propaganda in Germany, known how, by what methods, and for what purposes American analysts were examining his broadcasts during World War II, he would have found ways to deceive the analysts. Individuals can be taught how to achieve high scores on aptitude tests, and those who believe that success in their chosen career paths depends on their scoring well on these tests eagerly seek appropriate education. The extent to which preparatory instruction improves students' scores on a given test is also the extent of that test's invalidity. As an unobtrusive technique, content analysis can avoid such biases altogether.

Content analysis can handle unstructured matter as data. For efficiency's sake, researchers gain a considerable advantage if they can impose a structure on the data-making process so that the results are readily analyzable. Surveys, mail questionnaires, and structured interviews typically offer respondents predefined choices that are easily tabulated, coded, or processed by computer. But they thereby also prevent the respondents' individual voices from being heard. Subjects in laboratory experiments are often taught what amounts to a highly artificial data language: pushing buttons, scaling their opinions numerically, identifying shapes or forms they may never have seen before, or administering electric shocks to fellow subjects in place of less clearly measurable expressions of violence. These techniques are successful because they allow researchers to suppress unwieldy variations, which are due largely to the fact that ordinary human subjects see, talk, and behave in many different ways.

Typically, content analysts become interested in data only after the data have been generated. They have to cope with texts in a diversity of formats associated with different purposes, do not always find what they are looking for, and cannot fully anticipate the terms and categories used by the sources of their texts. This puts content analysts in an analytical position that is less than advantageous, a condition they share with ethnomethodologists, anthropologists doing fieldwork, historiographical researchers, and researchers using hermeneutical or interpretive approaches (such as those used in studies of politics, psychotherapy, feminist scholarship, and social constructionism). The chief advantage of the unstructuredness of content analysis data is that *it preserves the conceptions of the data's sources,* which structured methods largely ignore.

Content analysis is context sensitive and therefore allows the researcher to process as data texts that are significant, meaningful, informative, and even representational to others. Context-insensitive methods, such as controlled laboratory experiments, surveys, structured interviews, and statistical analyses, generate data without reference to their original contexts, thus disembodying observations, unitizing complex and contiguous events, and taking single words out of their contexts of use and representing them as data points in the analysts' theoretical spaces. In such methods, it no longer matters what gave rise to the data, how various elements in the data relate to each other, how others understand the data, or what the data mean to

their sources. Context-sensitive methods, in contrast, acknowledge the textuality of the data—that is, they recognize that the data are read by and make sense to others, and they proceed by reference to contexts of their own. Inferences drawn through the use of such methods have a better chance of being relevant to the users of the analyzed texts.

Content analysts may not always be as qualitative as are political analysts, who live in the very process they analyze. Nor are they quite as free as cultural studies researchers and conversation analysts, who contextualize their texts in a vocabulary that may appear alien to the people they speak for or of. Finally, content analysts may not be quite as limited in scope as the users of projective tests, who confine themselves to inferences concerning individuals' psychological characteristics (much as in content analyses of the second kind of definition discussed above).

Content analysis can cope with large volumes of data. Much of ethnomethodology as well as case study approaches, historiographical methods, and interpretive research relies on small samples of text; the volume of data is limited largely by what a researcher can read reliably and without losing track of relevant details. Although content analysis can be used to analyze small samples of texts (in fact, this is quite common, especially in the academic world, where funds are few and stakes are not as high as in politics, commerce, or medicine), such uses do not realize the technique's full potential. The ability to process large volumes of text in content analysis is paid for by the explicitness of the method's procedures, which, if clearly stated, can be applied repeatedly, by many coders or by computer software. As noted above, Berelson and Lazarsfeld (1948) stated long ago that content analysts must be systematic in their reading of texts and suggested that category schemes be devised that can be applied to every unit of text equally and without exception. Explicit vocabularies enable content analysts to employ many coders and pool their readings, which allows them to process quantities of text that far exceed what single individuals can analyze reliably. Consider the following numbers of units of analysis processed in early content analyses, largely without the aid of computers:

- 481 personal conversations (Landis & Burtt, 1924)
- 427 school textbooks (Pierce, 1930)
- 4,022 advertising slogans (Shuman, 1937; cited in Berelson, 1952)
- 8,039 newspaper editorials (Foster, 1938)
- 800 news of foreign language radio programs (Arnheim & Bayne, 1941)
- 19,553 editorials (Pool, 1952a)
- 15,000 characters in 1,000 hours of television fiction (Gerbner et al., 1979)

These numbers, which were considered impressive in 1980, when the first edition of *Content Analysis* was published, are now dwarfed by the sizes of electronic full-text databases that have emerged since. At the time of this writing, in the United States, ERIC, a clearinghouse for educational and social science writing, has more than 800,000 articles in its database. NewsBank, Inc., provides many libraries with Web-based access to current and archived text from more than 2,000 newspaper

titles. Factiva, founded jointly by Dow Jones and Reuters, collects documents from nearly 28,000 sources from around the world, in 23 languages, including 600 continuously updated news wires and more than 2,300 sources available on or before their date of publication. LexisNexis provides online access to billions of searchable documents and records from more than 45,000 legal, news, and business sources, some accumulating since the late 1980s. EBSCO MegaFILE is a textual database designed specifically for academic institutions; it provides full-text access to more than 4,600 journals and twice as many abstracts of work in many scholarly disciplines. The Vanderbilt News Archive specializes in television news and videos. Its core collection includes U.S. evening news broadcasts from ABC, CBS, and NBC since 1968, an hour per day of CNN since 1995, and Fox News since 2004. Google has started its project of scanning an estimated 129 million published books in the world, with more than 12 million books scanned as of June 2010.

In addition to traditional print media, the exponentially growing blogosphere, consisting of individually maintained internet blogs and their connections, is an unimaginably large source of content analysis data. The blogosphere constitutes a "community of texts," including social networks (Twitter and Facebook) and other exchanges. Rogers, Jansen, Stevenson, and Weltevrede (2010), for example, collected 653,883 tweets, tagged "iranelections RT," exchanged among Iranians in the period June 10–30, 2009, to analyze the chronology of changing concerns right after Iran's contested election. These are big numbers. The blogosphere is currently estimated to double in size every six months. The expanding use of social networks and the instantaneous availability of electronic texts have had the effect of bringing content analysis closer to large population surveys—less expensive, more comprehensive, yielding faster results, and being only minimally obtrusive. The relatively slow development of software to cope with such bodies of text is also shifting the bottleneck of content analysis from the traditional costs of access and tedious human coding to the need for good theory, sound methodology, and software that is capable of coping with such volumes. Here, pioneering work is needed and emerging.