1989

Content Analysis

Klaus Krippendorff
University of Pennsylvania, kkrippendorff@asc.upenn.edu

Content Analysis

CONTENT ANALYSIS

Content analysis is indigenous to communication research and is potentially one of the most important research techniques in the social sciences. It seeks to analyze data within a specific context in view of the meanings someone—a group or a culture—attributes to them. Communications, messages, and symbols differ from observable events, things, properties, or people in that they inform about something other than themselves; they reveal some properties of their distant producers or carriers, and they have cognitive consequences for their senders, their receivers, and the institutions in which their exchange is embedded. Whereas most social research techniques are concerned with observing stimuli and responses, describing manifest behaviors, differentiating individual characteristics, quantifying social conditions and testing hypotheses relating these, content analysis goes outside the immediately observable physical vehicles of communication and relies on their symbolic qualities to trace the antecedents, correlates, or consequences of communications, thus rendering the (unobserved) context of data analyzable. The methodologically critical requirement of any content analysis is to justify the inferential step this involves.

Definition. Formally, content analysis is a research technique for making replicable and valid inferences from data to their context. This definition encompasses those of Bernard Berelson, which equates content analysis with the scientific "description of the . . . content of communication," and of HAROLD D. LASSWELL, which emphasizes the quantification of the "what" that messages communicate. Ole R. Holsti adds such antecedents as the "who" (the source), the "why" (the encoding process), the "how" (the channel), and the consequences or "effects" they have "on whom." Although conventional conceptions of content (what) and communication contexts (who says it to whom) are common in content analy-

sis, the formal definition encompasses other communicative circumstances and contexts, such as psychoanalytical (psychological conditions that explain a particular statement), institutional (socioeconomic interests that underlie a particular television program), and cultural (functions that particular rituals serve).

Data for content analysis. The most obvious sources of data appropriate for content analysis are texts to which meanings are conventionally attributed: verbal discourse, written documents, and visual representations (*see* MEANING). The mass media have been the most prominent source, and the literature is dominated by content analyses of newspapers, magazines, books, RADIO broadcasts, films, comics, and television programming. However, the technique is increasingly applied to data that are less public: personal letters, children's talk, disarmament negotiations, witness accounts in courts, audiovisual records of therapeutic sessions, answers to open-ended interview questions, and computer conferences. Data that are meaningful only to small groups of experts are also considered: postage STAMPS, motifs on ancient pottery, speech disturbances, the wear and tear of books, and dreams. Anything that occurs in sufficient numbers and has reasonably stable meanings for a specific group of people may be subjected to content analysis.

Uses of Content Analysis

Content analysis rarely aims at a literal description of communications content. Examples of exceptions are efforts to determine whether a radio station used obscene language in its broadcasts or to establish the exact phrasing of a politician's campaign commitments. The ability to support inferences that go beyond the unaided understanding of a text is largely due to the systematic treatment of content analysis data. Ordinary readers (including literary scholars) tend to change their perspective as they read through large volumes of material and to be selective in support of their favored hypotheses. Content analysis assures not only that all units of analysis receive equal treatment, whether they are entered at the beginning or at the end of an analysis but also that the process is objective in that it does not matter who performs the analysis or where and when. Moreover, content analysis allows researchers to establish their own context for inquiry, thus opening the door to a rich repertoire of social-scientific constructs by which texts may become meaningful in ways that a culture may not be aware of. Both features enable the content analyst to provide aggregate accounts of inferences from large bodies of data that reveal trends, patterns, and differences no longer obvious to the untrained individual.

Studies of media content. Probably the most widespread use of content analysis is to infer the importance writers, producers, media, or even whole cultures assign to particular subject-matter categories from the frequency or volumes with which such subject matter is mentioned. Early examples are analyses of how the attention by newspapers to particular news categories has changed over time, how advertisements intrude into the coverage of religious matters, and how SPORTS and crime have taken over space from the cultural sphere. Others sought to explain differences among newspapers' attention in terms of ethnicity, readership characteristics, economics, and so on. Many of these studies are motivated by the feeling that journalistic standards are inadequately applied. For example, concerns for fairness are implied in numerous content analyses that aim to show the inequality of the coverage of the two (or more) sides of a public controversy, the imbalance in the favorable and unfavorable treatment of an issue, public figure, or foreign country.

The attention paid to particular phenomena, ideas, or ATTITUDES is the target of many social research efforts. An early content analysis showed how the image of popular heroes in MAGAZINE fiction had changed over a forty-year period from merchant-entrepreneurs to entertainers. Similarly, the images of teachers, scientists, police officers, and important politicians have been studied comparatively, in different media, and over time. A worldwide content analysis of the POLITICAL SYMBOLS in the prestige papers of several countries attempted to discern structural changes in governments and to predict revolutions. Analyses of the demographic, socioeconomic, ethnic, and professional characteristics of the population of television characters yielded considerable biases when compared with corresponding audience characteristics. Longitudinal studies of the kind, magnitude, and frequency of television VIOLENCE, attitudes toward war, the role women assume in popular serials, and arguments used to sell products or services have provided a basis for cultural criticism and have made the public aware of how the mass media may create particular beliefs or reinforce existing prejudices. Content analyses of news sources and references to foreign countries in various national media have demonstrated considerable imbalances in international news flow and attention. Systems for monitoring a corporation's symbolic environment by content analyzing newspaper clippings on issues and public attitudes of interest to that corporation have aided experiments in public relations and signaled significant changes in product perception, corporate image, state of the competition, and so forth.

Evidence provided by content analysis has been presented in U.S. courts in cases of plagiarism and COPYRIGHT violations and in a famous case involving

the identity of foreign news bureaus operating in the U.S. Inferences regarding the latter were based on various tests revealing privileged access to information, consistency with stated (foreign) PROPAGANDA aims, and deviations from neutral news sources.

Intelligence gathering and political studies. A government's knowledge about political developments in foreign countries often relies on communications in the form of diplomatic correspondence, foreign broadcasts, journalistic accounts in the domestic press, or speeches made by political leaders not necessarily intended to reveal these developments. Although political analysts typically do not take the time to codify their methods of drawing inferences from such data, there are several examples in which content analysis has provided important insights. Speeches by Soviet Politburo members delivered on the occasion of Joseph Stalin's birthday revealed the power structure within the Politburo and shed light on the anticipated succession to power (substantiated after Stalin's death). During World War II inferences about the war mood in Germany and changes in the relationships between Axis countries were based on systematically monitored domestic broadcasts. Similarly, the speeches of JOSEPH GOEBBELS, intended to boost German morale and prepare the population for forthcoming events, were successfully used to extract military intelligence. Content analyses to monitor a country's compliance with strategic arms limitation agreements have been proposed. Content analyses of communications exchanged on the eve of World War I, during the Cuban Missile Crisis (1962), and in the Sino-Soviet conflict (late 1950s) have employed interactional constructs, viewing the diplomatic and public statements made by leaders of countries in conflict as stimulus-response sequences.

Beyond these uses in politics, content analyses have shed light on the kind of values expressed and attitudes held on particular issues by candidates for political offices, and how these change in response to particular circumstances and with the kind of audiences addressed. Cross-national comparative analyses revealed differences in leadership values and elite aspirations, and studies of party platforms, British crown speeches, and Soviet May Day slogans established trends of interest to political scientists. Early efforts to detect propaganda techniques and to identify the "propagandists" thought to use these techniques to undermine rational judgments might be mentioned here as well.

Social sciences and literature. Much of the determination of individual psychology rests on verbal responses to interview questions, taped psychoanalytic sessions, diaries, essays, and letters (*see* DIARY; LETTER). The search for patterns of speech that would indicate particular psychopathologies is widespread. Content analysis is virtually built into projective tests, in which subjects give verbal responses to standardized stimuli that are then categorized and counted. Content analysis is also used to infer various psychological states of a speaker, such as the level of anxiety from the frequency of a speaker's speech disturbances (*see* SPEECH ANXIETY), or a speaker's idiosyncratic worldview from the kind of logical errors made during arguments. The kind of construct underlying the latter inferences has been applied to suicide notes, talk by alcoholics, and speeches by historical figures. Personal letters are similarly analyzed to reconstruct the individual dynamics and personality of the writer. Verbal records of dreams are a similar source of insights that content analysis opens to systematic inquiry.

In EDUCATION content analysis has been used to analyze textbooks for the sexual, racial, and national prejudices they contain, including how the depiction of wars differs in the history books of former enemies. An important educational use of content analysis is to infer the readability and reading interest of a text from the kinds of words, grammatical constructions, punctuation, and so on used. *See* READING; TEXTBOOK.

A natural candidate for content analysis is literature. The construction of concordances and the analysis of literary metaphors, symbols, themes, figures of expression, styles, GENRE differences, and intended audiences all fall into the domain of content analysis but are rarely so called. An interesting exception is the statistical identification of the unknown author of a literary work, successfully accomplished in the case of the medieval text *De Imitatione Christi,* several unsigned *Federalist* papers, and the differentiation of sections of a text written by different authors. The establishment of literary influences might be mentioned here as well as efforts to date documents by an analysis of WRITING styles and contents, attempts to infer achievement imagery in Greek literary works, and searches for themes that differentiate best-sellers from other novels (*see* STYLE, LITERARY).

Content analysis may be an integral part of a larger social research effort. For example, to minimize interviewer biases, open-ended answers to interview questions are often subjected to content analysis in order to obtain frequency distributions, scales, indexes, or variables that can then be correlated with directly measurable interviewee characteristics. Experiments with small (problem-solving, conflict simulation, or therapeutic) groups employ content analyses to differentiate among kinds of verbal interactions, to quantify the contributions made by members, and to conceptualize the role they assume in directing the emergence of social structures that may become explainable in these terms. MASS MEDIA EFFECTS have been studied by correlating content analysis measures of fictional television violence with estimates of ac-

tual violence obtained from heavy and light viewers.

Content analysis may also parallel other research techniques and check or shed light on the validity of either's findings. For example, a comparison of the actual crime statistics in a U.S. city, a poll of residents' concern about violence in that city, and a content analysis of the crime coverage in the local newspapers showed significant correlations only between the latter two, suggesting that they both indicate similar phenomena that are only marginally related to the "facts." A content analysis of essays written by students yielded findings substantially similar to those obtained from attitudinal questionnaires filled out by the same students. Such parallelisms enhance the analysts' confidence in the validity of their findings and justify their substitutability. On the other hand, an effort to find a strong correlation between various ways of counting references to U.S. presidents in a political science text, commonly accepted as a measure of attention or importance, and rank orders of the significance of these presidents provided by the author of that text failed to reach acceptable levels. This speaks against the unreflected use of frequency measures and points to the importance of establishing the validity of any content analysis.

Procedures and Their Criteria

Content analyses commonly contain six steps that define the technique procedurally.

Design. A conceptual phase during which analysts define their *context,* what they wish to know and are unable to observe directly; explore the source of relevant data that either are or may become available; and adopt an analytical construct that formalizes the knowledge available about the data-context relationship, thereby justifying the inferential step involved in going from one to the other. These three principal features constitute the framework for analysis (see Figure 1). Besides delineating the empirical procedures to be employed, the design should also spell out the observational conditions under which the inferences made could be considered valid, in the sense of representing what they claim to represent.

Unitizing. The phase of defining and ultimately identifying units of analysis in the volume of available data. *Sampling units* make possible the drawing of a statistically representative sample from a population of potentially available data: issues of a newspaper, whole books, television episodes, fictional characters, essays, advertisements. *Recording units* are regarded as having meanings independent of one another: references to events, individuals, or countries; evaluative assertions; propositions; themes.

Sampling. While the process of drawing representative samples is not indigenous to content analysis, there is the need (1) to undo the statistical biases inherent in much of the symbolic material analyzed (e.g., the attitudes of important people are expressed more frequently in the mass media than are those of the larger population) and (2) to ensure that the often conditional hierarchy of chosen sampling units (e.g., publications–newspaper dates–page numbers–articles–paragraphs–words) becomes representative of the organization of the symbolic phenomena under investigation.

Coding. The step of describing the recording units or classifying them in terms of the categories of the

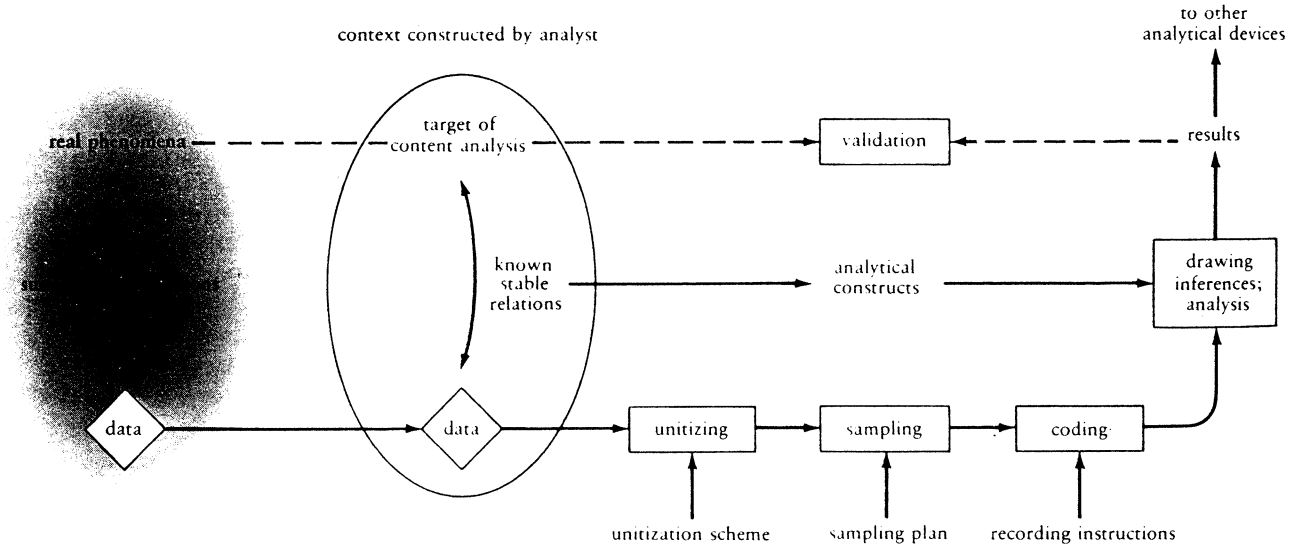

Figure 1. *(Content Analysis)* The content analysis research process. Diagram by Klaus Krippendorff.

analytical constructs chosen. This step replicates an elementary notion of meaning and can be accomplished either by explicit instructions to trained human coders or by computer coding. The two evaluative criteria, reliability as measured by intercoder agreement and relevance or meaningfulness, are often at odds. Human coders tend to be unreliable but good at interpreting semantically complex texts (*see* INTERPRETATION). Computers have no problems with reliability but must be programmed to simulate much of a native speaker's linguistic competence. Notwithstanding major advances in the use of computers, their application usually sacrifices the criterion of meaningfulness in favor of reliability and speed.

Drawing inferences. The most important phase in a content analysis. It applies the stable knowledge about how the variable accounts of coded data are related to the phenomena the researcher wants to know about. The inferential step involved is rarely obvious. How the frequency of references indicates the attention a source pays to what it refers to, which distinct literary style uniquely identifies a particular author, and the way preferences for certain verbal attributions manifest speaker or listener attitudes need to be established by independent means. Analytical constructs of this kind need not be so simple either. In extracting military intelligence from enemy broadcasts, analysts employ elaborate "maps" of known relationships involving the role of and conflicts within the national leadership and among the population addressed. Similarly, inferences about individuals' worldviews from their idiosyncratic styles of reasoning involve several levels, each employing elaborate psychological constructs of their cognition.

Validation. The desideratum of any research effort. However, validation of content analysis results is limited by the intention of the technique to infer what cannot be observed directly and for which validating evidence is not readily available. For example, why would one want to extract military intelligence from enemy propaganda if the adversary's planned activities were already known, why would one want to infer media attention if attention were measurable directly, or why would one want to infer Kennedy's changing attitudes during the Cuban Missile Crisis from his communications if it were possible to interview him? Nevertheless, content analysis should not be undertaken without at least the possibility of bringing validating evidence to bear on its findings.

Limitations

Despite its claim to generality, content analysis has some inherent limitations. The first stems from its commitment to scientific decision making. Statistically significant findings require many units of analysis, and seeking such findings amounts to a commitment to be quantitative. This discourages the analysis of unique communications or connected (nondecomposable) discourses characteristic of literary, historical, or psychoanalytic inquiries.

The second limitation stems from the replicability requirement. This implies fixed and observer-independent categories and procedures that must be codified without reference to the analyst and the material being analyzed. Computer content analysis is one of its results. It favors the use of data in contexts that entail stable and unambiguous interpretations and leaves little room for those whose meanings evolve in the process of communication and in ways characteristic of the different communicators or social groups involved. Such ambiguities are frequent in political and private discourses.

The expectation to contribute to social theory leads to the third limitation. If categories are obtained from the very material being analyzed, findings are not generalizable much beyond the given data. If they are derived from a general theory, findings tend to ignore much of the symbolic richness and uniqueness of the data in hand. The compromises content analysis must seek are rarely easy ones.

See also AGENDA-SETTING; COMMUNICATIONS RESEARCH: ORIGINS AND DEVELOPMENT; MASS COMMUNICATIONS RESEARCH.

Bibliography. Bernard Berelson, *Content Analysis in Communication Research*, Glencoe, Ill., 1952, reprint New York, 1971; George Gerbner, Ole R. Holsti, Klaus Krippendorff, William J. Paisley, and Philip J. Stone, *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*, New York, 1969; Ole R. Holsti, *Content Analysis for the Social Sciences and Humanities*, Reading, Mass., 1969; Klaus Krippendorff, *Content Analysis: An Introduction to Its Methodology*, Beverly Hills, Calif., 1980; Harold D. Lasswell, Nathan Leites et al., eds., *Language of Politics*, Cambridge, Mass., 1949; Harold D. Lasswell, Daniel Lerner, and Ithiel de Sola Pool, *The Comparative Study of Symbols*, Stanford, Calif., 1952; Klaus Merten, *Inhaltsanalyse*, Opladen, FRG, 1983.

KLAUS KRIPPENDORFF

International Encyclopedia of Communications

ERIK BARNOUW
Editor in Chief

GEORGE GERBNER
Chair, Editorial Board

WILBUR SCHRAMM
Consulting Editor

TOBIA L. WORTH
Editorial Director

LARRY GROSS
Associate Editor

Volume 1

Published jointly with
THE ANNENBERG SCHOOL OF COMMUNICATIONS,
University of Pennsylvania

OXFORD UNIVERSITY PRESS
New York Oxford