CHAPTER 2

Milestones in the History of Content Analysis

This chapter examines the motley history of content analysis and identifies prominent projects and trends that have contributed to enormous growth in the use of the method. Key players are identified in the development of content analysis as the prominent technique it is today.

The Growing Popularity of Content Analysis

Content analysis has a long history of use in communication, journalism, sociology, psychology, and business. Content analysis is being used with increasing frequency by a growing array of researchers. Riffe and Freitag (1997) note a nearly six-fold increase in the number of content analyses published in *Journalism & Mass Communication Quarterly* over a 24-year period—from 6.3% of all articles in 1971 to 34.8% in 1995, making this journal one of the primary outlets for content analyses of mass media. And the method of content analysis is more frequently taught at universities. By the mid-1980s, over 84% of master's-level research methods courses in journalism included coverage of content analysis (Fowler, 1986).

The growth in the use of content analysis over the past four decades may be seen in a simple analysis shown in Box 2.1, with highlights of important trends graphed in Figure 2.1. Box 2.1 contains tabled frequencies for library online index and abstract searches for the terms *content analysis* and *text analysis*. The indexes and abstracts were chosen for their ready availability and their potential relevance to content analysis in a social or behavioral science context.[1] The 20 indexes and abstracts used in the analysis cover a wide range of interests and publication types.[2]

Box 2.1 Content Analysis Timeline

Search for Keyword *Content Analysis* or *Text Analysis*

Index or Abstract	*Begins*	*Total Number*	*1958*	*1959*	*1960*	*1961*	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*	*1977*
ABI/Inform	1971	482														0	0	0	1	6	2	9
Anthropological Literature	1900	65	0	1	1	0	0	1	0	0	2	0	0	0	0	0	2	3	0	1	1	4
Arts & Humanities	1980	83																				
ComAbstracts	1966	265									1	1	0	0	0	1	0	0	0	0	1	1
Communication Abstracts	1998	292																				
Educational Abstracts	1983	1088																				
ERIC	1966	6022									12	27	58	111	103	121	160	176	198	192	236	207
HealthSTAR	1975	1305																		12	12	11
Historical Abstracts	1960	135	2	0	1	0	0	0	0	2	0	1	0	0	4	4	2	9	2	2	5	6
Humanities Abstracts	1984	213																				
Library Literature	1984	244																				
Medline	1966	1507								1	2	2	3	6	4	4	7	5	6	12	17	3
Music Index	1979	44																				
PAIS	1972	126															0	0	0	0	2	4
Periodical Abstracts	1986	404																				
ProQuest Digital Dissertations	1861	5832	4	5	6	8	3	5	6	9	12	12	10	17	25	23	18	25	29	26	23	19
PsychInfo	1887	4403	8	13	15	4	18	12	14	16	24	54	34	33	17	34	38	47	67	57	55	80
Social Sciences Cite Index	1980	1839																				
Sociological Abstracts	1963	3630						7	9	8	7	10	14	10	17	23	40	37	29	57	36	90
World Cat	1200	4225	11	25	22	34	21	32	18	37	48	64	61	62	98	85	101	90	92	98	98	126

Index or Abstract	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*
ABI/Inform	2	9	11	9	13	15	22	14	19	23	22	21	27	17	36	42	38	33	28	29	13	24
Anthropological Literature	4	0	8	12	6	2	2	1	2	1	1	1	0	1	2	0	1	1	1	3	1	1
Arts & Humanities			1	5	2	1	5	1	2	7	3	4	9	12	19	17	15	15	14	9	10	3
ComAbstracts	1	2	0	1	0	2	1	0	3	1	4	5	27	29	11	17	22	19	29	24	26	31
Communication Abstracts																					145	147
Educational Abstracts						45	52	80	71	56	63	54	60	50	61	74	61	72	68	84	52	84
ERIC	211	229	238	235	233	230	230	276	251	225	202	204	185	159	189	199	202	205	239	226	183	129
HealthSTAR	11	10	10	10	20	20	11	29	24	16	35	55	78	82	76	88	106	98	139	126	129	173
Historical Abstracts		6	4	7	5	4	11	2	6	7	8	7	6	8	2	2	3	9	3	0	2	2
Humanities Abstracts							1	3	7	6	5	8	15	15	21	23	29	21	26	22	17	16
Library Literature							1	2	1	11	14	16	26	21	24	16	24	15	20	25	21	25
Medline	15	13	2	20	25	25	16	39	29	23	45	63	100	96	84	98	126	104	152	141	130	172
Music Index		3	3	3	6	4	4	5	4	8	7	8	11	6	7	5	12	8	8	9	3	13
PAIS	3	0	2	3	2	3	5	5	2	14	4	4	11	8	7	2	7	4	5	13	10	17
Periodical Abstracts									5	0	3	35	20	16	37	47	43	38	50	58	28	64
ProQuest Digital Dissertations	18	55	202	228	203	214	223	226	238	264	263	284	275	321	341	357	359	416	346	382	345	381
PsychInfo	67	53	93	84	146	108	94	147	166	138	155	157	189	207	211	249	244	257	289	327	279	310
Social Sciences Cite Index			20	15	27	29	21	23	21	28	34	31	30	55	149	153	163	183	201	266	217	275
Sociological Abstracts	88	65	56	58	75	68	90	135	202	178	160	160	179	184	192	201	256	272	287	211	125	122
World Cat	101	131	133	112	128	130	120	145	129	160	126	156	143	162	189	184	176	155	175	171	157	116

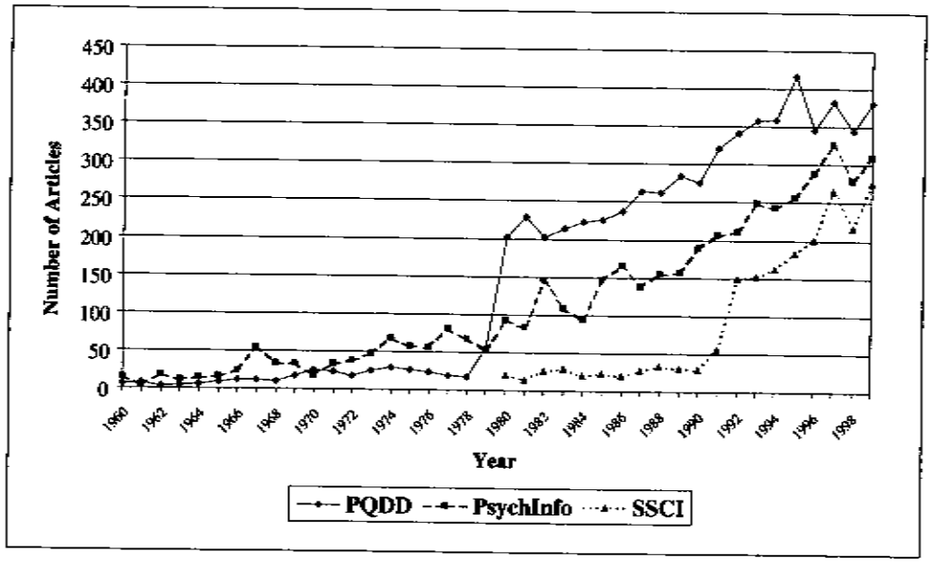

Figure 2.1. Timeline: Content Analysis and Text Analysis Articles by Year

The tabled numbers should be viewed cautiously and interpreted for what they are—the results of simple searches for terms, in publications available since 1958[3]— without full information about how the terms have been used by the researchers. That is, a number of so-called content analysis studies are actually qualitative studies, not what this book would define as content analysis at all. Second, the indexes overlap in their coverage. For example, a number of communication journals are indexed in both Communication Abstracts and PsychInfo. Third, some of the growth in content analysis applications shown is surely due to the expansion in the number of journals indexed (via new journals and cross-over additions). Taking these caveats into account, the evidence is still clear: Never has content analysis received more attention in the research literature than at the present time.

The expanding use of the technique is particularly notable in the searches of PsychInfo, the Social Science Citation Index, and ProQuest Digital Dissertations; results for searches of these three indexes are graphed in Figure 2.1. The growing use of content analysis as a technique for graduate theses and dissertations is particularly noteworthy. This expansion is shown in the rising numbers both for ProQuest Digital Dissertations and for World Cat (which includes many master's theses).

The history of the use of the various techniques called content analysis has been documented only piecemeal, with some histories emphasizing text analysis (e.g., Stone et al., 1966); some, computer text analysis (e.g., Diefenbach, in press); and others, specific applications (e.g., Rosengren, 1981), including communication (Berelson, 1952) and psychological diagnosis (Gottschalk, 1995). The field is in need of a comprehensive history that melds the various cross-influences, which are many. Unfortunately, this book, with its more practical bent, does not allow room for a comprehensive history. Instead, 10

milestones important to the history of content analysis have been selected, with an eye to those events, persons, institutions, and trends whose influence has transcended disciplinary and substantive boundaries. Following the model of Lowery and DeFleur (1995), both key historical details and important implications for the advancement of knowledge are given for each featured milestone.

Milestones of Content Analysis Research

Rhetorical Analysis

The first recorded message analyses were not content analyses. They were subjective, normative essays consisting of advice on effective speaking, and they were written about 4,000 years ago (McCroskey, 1993). In the beginning of message analysis, Aristotle applied his considerable energies to the study of rhetoric, the "art of discovering persuasive proofs" (Miller, 1987, p. 447). His tripartite analysis, focusing on the speaker, the audience, and the speech, put message content and form at the center of an argument—that we use communication to control our environment, including the actions of others (Aristotle, 1991). Over the subsequent millennia, rhetoricians have continued to conduct microscopic and theory-based analyses of persuasive communication. While rhetorical scholars prefer a logical, qualitative approach to message analysis, their techniques complement the quantitative approaches of social and behavioral scientists (McCroskey, 1993). And quite simply, we wouldn't be doing what we do if not for the influences of the early rhetoricians, who first took the art and practice of human communication seriously.

Biblical Concordances and the Quantification of History

Content analysis "learned its methods from cryptography, from the subject classification of library books, and from biblical concordances, as well as from standard guides to legal precedents" (Marvick, as quoted in Rogers, 1994, p. 214). Such systematic indexing of large message sets brought order and retrievability to the study of massive historical documents. Concordances, systems of cross-listings of terms and concepts so that themes might be readily accessed in a large text, were used centuries ago to organize scholarly studies of the Bible. Such concordances have been of use in organizing other historical documents, such as the Dead Sea Scrolls. (Obviously, the advent of the computer has made such tasks much easier.)

Another trend in the quantification of history is decryption, the identification of intended meanings for written symbols. The Rosetta Stone, discovered in 1799, contained three parallel script types—hieroglyphs, demotics, and Greek letters. Thomas Young was able to "translate" between the three scripts through a process of quantifying (counting) occurrences of signs on the stone and other ancient sources.

Other historians have applied statistical procedures to the task of organizing data from historical documents. The massive effort by Robert Fogel and Stanley Engerman (1974a, 1974b) to study American slave treatment, *Time on the Cross,* attracted both popular media and scholarly attention. The comprehensive analysis was based on historical documents containing pieces of evidence that included birth records; slave sale prices; the dimensions of slave cabins; and many measures of nutrition, health, and family life extracted from plantation records. Their analyses culminated in a report that concluded that American slaves enjoyed greater physical and psychological well-being than had previously been supposed (Sutch, 1975). Controversial in its conclusions from the date of publication (Sutch, 1975), *Time on the Cross* nevertheless succeeded in focusing attention on the meaning of "messages" sent, as it were, over time, and it stimulated interest in *cliometrics,* the quantitative measurement of historical events and trends.

While quantitative studies in history continue to be a viable part of the discipline (Dollar & Jensen, 1971; Jarausch & Hardy, 1991; Rowney & Graham, 1969), content analysis has not played as major a role in the field as have the more conventional techniques of historians (Carney, 1971; Floud, 1977). Still, content-analytic examples do exist in studies of history textbooks (e.g., Gordy & Pritchard, 1995; Holt, 1995), political documents (e.g., Anheier, Neidhart, & Vortkamp, 1998; Beriker & Druckman, 1996), news coverage of historical events (e.g., Switzer, 1990), and in other novel approaches capturing the tone of historical periods (e.g., Padilla's [1997] study of U.S. Army recruiting posters beginning in 1915 and Matcha's [1994-1995] study of early-20th-century obituaries in Ohio).

The Payne Fund Studies

In the days before television ruled the American scene, the most popular and compelling mass medium was narrative film. And before people were worried about the effects of violent television, they worried about the corruption of the nation's youth by movies. In the early years of sound films, the first major research efforts were made, backed with money from the Payne Fund. The studies looked at both movie content and effects, forging advancements in various methods of media study. Box 2.2 focuses on this milestone of an era "when movies were king."

The Language of Politics

The title of this subsection is appropriately taken from the name of Lasswell, Leites, and Associates' (1949) book, which explored the power of propaganda and the quantitative methods appropriate for examining the content of political messages. Harold Lasswell is very possibly the single most influential figure in the development of the systematic study of messages in the

Box 2.2 When Movies Were King

Studying the Power of the Moving Image

In the late 1920s, virtually everybody in America went to the movies regularly . . . and that means *everybody*—whole families attended, unlike today's teen and date audiences. Children above 7 years of age attended movies an average of once a week (Charters, 1933). In this environment of common exposure to powerful, vivid, fictional representations, there was for the first time widespread public concern over the impact of negative images and behaviors on children. In 1928, William H. Short, Executive Director of the Motion Picture Research Council, coordinated the establishment of a major research project, funded by The Payne Fund. The researchers were psychologists, sociologists, and educators from the University of Chicago, the University of Iowa, Yale University, New York University, Pennsylvania State College, Columbia University, and Ohio State University. They launched what would be the most ambitious research project on the effects of entertainment media to date and for several decades to follow.

The Payne Fund Studies have been identified as one of the landmarks of early mass communication research (Lowery & DeFleur, 1995). Through experimental and response studies, the movies' effects on children's learning and attitudes were studied. Surveys and interviews measured correlates of movie attendance among U.S. youth. Autobiographical case studies and experiments that employed galvanic skin response measures looked at the children's emotions as related to movie exposure. Even the effect of movies on children's sleep patterns was studied. The conclusions reached were that (a) movies were a potent source of education for children, and (b) although clearly an influence on children, movies were one among many influences that molded the experience of kids; the situation was complicated and in need of further study (Charters, 1933).

Ohio State University was the home of the content analysis portion of the Studies. There, Edgar Dale (1935) conducted an analysis of the major themes from the written descriptions of 1,500 movies released in 1920, 1925, and 1930.[4] The films were classified into 10 types. Although the coding scheme for this variable was emergent, that is, devised from the film descriptions themselves rather than set in advance, good scientific procedures were followed when Dale used 300 films for a reliability check of the final 10-category measure, finding an average percent agreement of 87%. The results indicated a strong prevalence of both love and crime, with substantial representation by films emphasizing sex and comedy. The percentage figures for the 3 years were as follows:

	1920	*1925*	*1930*
Love	45	33	30
Crime	24	30	27
Sex	13	17	15
Comedy	12	13	16
Other	6	7	12
	100	100	100

(continued)

Box 2.2 Continued

Although this large-scale analysis gave the researchers important information about what movie content children were being exposed to, for Dale, this was just the beginning. He trained coders and sent them to theaters to watch 115 movies. They were armed with a coding guide that asked for information about nine major "social values" areas, with numerous variables under each one (e.g., nature of American life and characters; nature of foreign life and characters; motivation of characters; emotional appeals to audience; crime, delinquency, and violence; relations of sexes; military situations; depiction of underprivileged peoples; deportment). In the age before video, such real-time coding in the movie theater was limited to current releases and required that the coders sit "near a light . . . to make satisfactory notes" (Dale, 1935, p. 5). Given these challenges, it's understandable that a less-than-optimal coding protocol was established (i.e., the coders read reviews of the films before viewing to prime themselves for major themes, and the coding was what we might generously call "open-ended"—coders were to write as much as they could during the film and then continue their narratives shortly after viewing the film). On the other hand, Dale continued to attempt to apply rigorous standards. He trained the coders in contemporary standards for psychological observation and reporting and established independent reliability by accepting only those details reported by more than one coder. Moreover, he conducted an even more in-depth analysis of 40 of the films by obtaining dialogue scripts and sending stenographers to the movies to supplement the dialogue with descriptions of settings, characters, and actions. Again, he included reliability assessment. Throughout, when Dale was unable to achieve high standards, he was forthcoming about the study's limitations.

The numerous and diverse findings were summarized by Dale (1935) in a "balance sheet for motion picture content" (p. 229), which included the identification of the following emphases in 1930 movies: portrayals of life in the upper economic strata, rather than life among the middle and lower economic strata; problems of the unmarried and young, rather than problems of the married, middle aged, and old; problems of love, sex, and crime, rather than other problems of everyday life; motifs of escape and entertainment, rather than motifs of education and social enlightenment; individual and personal goals, rather than social goals; physical beauty, rather than beauty of character.

It's amazing how similar the findings are to those of contemporary analyses and criticisms of television and movies. Dale put his finger on the pulse of popular entertainment, and it still seems to be beating the same beat.

20th century. It was he who defined communication as "*who* says *what* to *whom* via what *channel* with what *effect?*"[5] He spent decades exploring the *whats* as well as the *effects* of the messages, with a focus on politically motivated communication.

Although trained as a political scientist, Harold Lasswell has been called the "da Vinci of the behavioral sciences" (Smith, 1969, p. 41) for his productivity and his renaissance-like interest and proficiency in a multitude of areas of

study. He was "rambunctiously devoted to breaking down the man-made barriers between the social studies, and so acquainting each with the rest" (American Council of Learned Societies as quoted in Rogers, 1994, p. 203). He worked with linguists, anthropologists, law scholars, psychologists, and psychiatrists[6] on research that ranged from political behavior to psychoanalysis to culture (e.g., Lasswell, 1935; Rogow, 1969). He made an indelible impression on everyone who worked with him (Schramm, 1997).

Among all Lasswell's contributions, a standout is his influence on the development of content analysis as a quantitative technique. What's particularly impressive is his honest assessment of the weaknesses in his technique as his methods developed. In his progression, we can see the evolution of social and behavioral science methods in general and content analysis methods in particular. He called his 1927 dissertation a "content analysis" of World War I propaganda techniques, but it was actually very qualitative and critical (Rogers, 1994, p. 213). However, it *was* an empirical study and a first attempt at a classification of propaganda. (He studied leaflets dropped from balloons and airplanes and examined military recruitment posters.) Later, he was pointedly critical of his own anecdotal evidentiary style, noting that he didn't even report the "criteria for selection" of examples of the various propaganda styles he identified (Lasswell et al., 1949, pp. 41-42). Also, Lasswell proceeded to praise much-earlier endeavors for their "systematic definition and historical detail" of measurement (Lasswell et al., 1949, p. 44)—such studies as George Carslake Thompson's analysis of public opinion in the British population of the 1870s and several studies of newspaper coverage that used good representative sampling, topic coding, and column inch measures of coverage as early as 1900.

Lasswell made improvements to his own content analysis work and completed many quantitative studies of propaganda before and during World War II. During the war, he was chief of the Experimental Division for the Study of War-Time Communications in the U.S. Library of Congress, funded by the Rockefeller Foundation (Lasswell et al., 1949; Rogers, 1994). Later, he headed research efforts by Revolution and the Development of International Relations, a project at the Hoover Institute and Library on War, Revolution and Peace at Stanford University (Lasswell, Lerner, & Pool, 1952). Later in his career, motivated certainly by his wartime efforts to describe and predict Nazi communication activity, he seemed particularly enthusiastic about the possibility of discerning the Communist leanings of writers of pamphlets and other publications.[7]

Most of the lessons Lasswell learned in his lifetime about the utility of scientifically sound content analysis have been passed down to us. Lasswell declared that

> content analysis operates on the view that verbal behavior is a form of human behavior, that the flow of symbols is a part of the flow of events, and that the communication process is an aspect of the historical process. . . .

> Content analysis is a technique which aims at describing, with optimum objectivity, precision, and generality, what is said on a given subject in a given place at a given time. (Lasswell et al., 1952, p. 34)

He devised methods of pilot testing, coder training, and reliability assessment that have served as early models of current sound practice.

The study of political propaganda was advanced by other researchers as well and has continued to this day. Coming out of the Institute for Propaganda Analysis (founded in 1937 with social psychologist Hadley Cantril at the helm), *The Fine Art of Propaganda* (Lee & Lee, 1939) presented seven common devices of propaganda[8] still in use today. In recent decades, the Manifesto Research Group, established by David Robertson and Ian Budge in 1979 and now working in cooperation with the Research Unit in Institutions and Social Change of the Science Centre-Berlin (Budge & Hofferbert, 1996), has systematically content analyzed party platforms for elections in many countries, going as far back as 1922. An archive of their data from 20 nations covering the period from roughly 1945 to 1983 is housed at the University of Essex. In their analyses, the Manifesto group has pinpointed clearly contrasting emphases of competing parties; for example, in pooled analyses for national elections in Great Britain from 1945 through 1992, they found the Conservative Party to emphasize free enterprise and law and order much more so than the Labour Party, which emphasized economic planning and social justice. In analyses of the U.S. national elections from 1948 through 1992, they found the Republican Party focusing on a positive portrayal of the military, government effectiveness, and economic orthodoxy more than did the Democratic Party, which focused on peace and social justice (Budge & Hofferbert, 1996, p. 86).

The War at Home: Advances in Social and Behavioral Science Methods During World War II

As noted, Lasswell and others developed content-analytic techniques for uncovering enemy orientations that were useful during World War II (Rogers, 1994, p. 224). This was just one of a host of social and behavioral science advances spurred on by the war effort. The period was marked by unprecedented cooperation between private industry, government, and scholarly research bodies, joined for the noble cause of winning "the good war." The discoveries and advancements of this era jump-started research initiatives in all the sciences, including the social and behavioral sciences. Some have noted that advances during this time gave the social sciences a measure of acceptance they had long sought (Delia, 1987).

The U.S. War Department's Information and Education Division conducted a series of experiments on the effectiveness of some of the seven-part *Why We Fight* series of training and indoctrination films produced for the military by Hollywood director Frank Capra. Although the films tested had an impact on the acquisition of factual knowledge, they did not result in the hoped-for morale boost (the type of patriotism enhancement that is assumed to have resulted from the films of Nazi documentarist Leni Riefenstahl[9]). The film evaluation studies set new standards for communication effects research and established the importance of individual differences in audience members' responses to the persuasive messages (Lowery & DeFleur, 1995). Also, the captive population of military personnel afforded social scientists a great opportunity to apply theories of sampling and to refine measurement procedures with individuals as the unit of data collection (Babbie, 1995).

As Rogers (1994) notes, "The banks of the Potomac were an exciting place for social scientists during World War II" (p. 224). In the midst of this activity, content analysis experienced its share of advancements. Code breaking and other surveillance activities (what today we would call "intelligence activities") demanded the application of linguistic and psychological theories to practical predictions. During World War II, a group "connected with the BBC" (Lasswell et al., 1949, p. 49) systematically analyzed radio broadcasts from Axis powers and began forecasting Nazi policy and activities in real time. And Allied forces were able to estimate the concentration of German troops in various locations by comparing music played on German radio stations with music played elsewhere in occupied Europe (Wimmer & Dominick, 1994, p. 163).

As Delia (1987) notes, the wartime advances in research were solidified and guaranteed continuation by an "unparalleled expansion of American higher education in the postwar period. As educational opportunities were opened through government support, the universities became engines of social and economic transformation" (p. 56).

Speech as a Personality Trait

Taken from Edward Sapir's (1927) important essay, the title of this section denotes the contribution of a host of linguistic and psychological studies that took to heart Sapir's admonition that human personality traits might be discerned from communication content.

> If we make a level-to-level analysis of the speech of an individual and if we carefully see each of these levels in its social perspective, we obtain a valuable lever for psychiatric work. . . . If carried far enough, [it] may enable us to arrive at certain very pertinent conclusions regarding personality.[10] (p. 905)

As detailed in Chapter 9, a number of diverse yet overlapping areas of content-analytic study have grown up around this approach. In the following discussion, some of their common roots are considered.

The use of content analysis for both basic psychometrics (i.e., the measurement of psychological traits and states) and clinical diagnoses owes a great deal to the work of Gordon Allport (1942, 1965), who, although not engag-

ing in content analysis per se, advanced the notion that the systematic inspection of personal documents (e.g., letters, diaries, open-ended responses on questionnaires) might be used for psychological measurement and in fact might shed light on theories of personality. In his editing and interpretation of the classic *Letters from Jenny* (Allport, 1965), he further posed the question of whether the study of a single case may fit the requirements of the scientific method. He argued that if the goals of science are to understand, predict, and control events, then to fulfill these goals for a single individual, we need to study the individual's own patterns of behavior and communication.

By the 1950s, true content analysis coding of verbalizations and individual texts was being used for a variety of psychotherapeutic purposes. Auld and Murray (1955), in their review of literature to that time, identified some of the more influential analysis schemes. Bales's (1950) interaction process analysis has been applied to exchanges between the counselor and the patient. Dollard and Mowrer's (1947) discomfort-relief quotient is a word-by-word content analysis system intended to measure the tension experienced by a patient involved in psychotherapy. Raimy's (1948) positive-negative-ambivalent quotient is another scheme intended to measure the patient's emotional reactions to the counseling process. These and other researchers established content analysis as a viable method for tapping individual traits and states.

Another innovative contributor to progress in content-analytic measurement of psychological constructs was David McClelland (1984), who developed a system of coding Thematic Apperception Test responses (from the work of Murray [1943]) to measure an individual's achievement motive (Winter, 1998). McClelland (1984) favored the development of standard, systematic content analysis coding schemes and was mystified by the field of psychology's initially slow progress in this regard: "perhaps it is the unusual sensitivity of the method to disturbances . . . [or] the time and trouble it takes to learn a coding system" (pp. 449-450). He compared psychology's reluctance to conditions in the natural sciences. "I knew that biologists were extremely careful in standardizing conditions for taking a measurement and would also spend hours making a single assay. Why couldn't psychologists take the same amount of trouble?" (p. 450).

By the 1950s, there was growing attention to methods of developing message-based personality measures from linguistic, historical, and emotional perspectives (Pool, 1959). Participants at the 1955 Allerton House conference at the University of Illinois came together to consider these and other issues. By the 1990s, psychologists' early reluctance had decreased to the point that dozens of standard content-analytic measures were in use (Smith, 1992).

Department of Social Relations at Harvard

Harvard during the 1960s was a center of activity on the study of human communication behavior, with the establishment of the General Inquirer

project. Philip Stone headed an interdisciplinary group that set out to use computer text analysis to content analyze written messages from linguistic, psychological, sociological, anthropological, and communication standpoints (Stone et al., 1966). With initial input from Harvard scholars Robert Bales and George Miller and supported by grants from the National Institute of Mental Health and the National Science Foundation, the project developed the first computer program designed to categorize language. The main dictionary was the Harvard Third Psychosociological Dictionary, which coded 3,564 different target words with 83 different tags (such as job role, clothing, and body part). As of 1965, Stone et al. noted 16 additional dictionary systems that had been developed for use with the General Inquirer, including the Stanford Political Dictionary, the Davis Alcohol Dictionary, two need affiliation dictionaries, and the Lasswell Value Dictionary (pp. 140-141). The program, with a high degree of flexibility to add and refine dictionary search and coding commands, is still actively used today, and the Essex Summer School features seminars in its use (see *The Content Analysis Guidebook Online* for links). And as Diefenbach (in press) notes, the system "still serves as the model for the method" of computer text content analysis (p. 14).

Television Images: Violence and Beyond

Just as film was the medium of choice in the 1920s and 1930s, television took over the hearts and minds of Americans in the 1950s. Perhaps for the first time since the Payne Fund studies, there was interest enough in the effects of entertainment media content to launch large-scale projects with big-time backing. The primary focus was violence on television, and numerous researchers would spend the next several decades documenting its presence and effects. Others would look at the portrayals of women, minorities, or the elderly. Some would even look at prosocial content on television (Greenberg, 1980; see also examples in Chapter 9).

Whereas many individual researchers looked at specific aspects of violence, the largest and most long-term effort was by George Gerbner and his Cultural Indicators Project team. Their studies on the frequency and types of violence on U.S. entertainment television began in 1967, with a grant from the National Commission on the Causes and Prevention of Violence, and has continued in some form ever since, funded by a host of governmental, nonprofit, and private institutions.[11] It has continually found levels of TV violence that far outstrip measures of violence in real life. And although the studies of Gerbner's team and others have supported the notion that TV violence does play a role in creating and perpetuating images of a violent and scary world, the upshot has been limited to simple debate and the institution of content warning systems. Contemporary studies of violence on TV continue to find that little has changed (National Television Violence Study, 1997). On the other hand, the high-profile Cultural Indicators Project has provided an invigorating model for generations of content analysts to come.

In the context of these projects, researchers honed their skills in analyzing the *moving image*. They had to move beyond simple text and develop new methods that considered nonverbal and other behaviors, artifacts (e.g., clothing), and even production techniques.

The Power of Computing

The General Inquirer, of course, demonstrated the use of computers to automatically analyze text messages, but its use was dependent on access to a powerful mainframe and some level of sophistication in computer use. And, the first demonstration of computer text analysis at a communication conference, by Rick Holmes and Joe Woelfel in 1982, was achieved via a 300-baud phone line connection of a "dumb" terminal at the conference with the UNIVAC mainframe at the University of Albany (Joseph Woelfel, personal communication, June 2000). Their demonstration of CATPAC heralded a changing research environment.

With the diffusion of the personal computer, both text analysis and data analysis have been brought to the desktop. PCs have revolutionized the means by which messages are created, stored and archived, disseminated, and organized. (Chapter 4 deals in greater detail with message management via computer, and Chapter 6 clarifies the important role that computer text analysis currently plays; see also Diefenbach, in press.) Huge databases of speech transcriptions or mass media messages can be analyzed at a keystroke. The availability of sophisticated data analysis techniques (for *all* purposes, not just for content analysis) is at an all-time high. Cluster analysis and multidimensional scaling are no longer the exclusive province of "techies"—CATPAC is now an extremely user-friendly, nicely designed PC application.

But we must remember that the notion of the completely "automatic" content analysis via computer is a chimera. As Diefenbach (in press) notes, "The maxim still holds, 'garbage in, garbage out' " (p. 5). The human contribution to content analysis is still paramount.

The Global Content Analysis Village

One of the many global villages[12] in evidence today is the community of scholars and practitioners of content analysis who communicate via the Internet. In past decades, there have been several efforts to reduce schisms and to cross-fertilize interdisciplinary approaches to content analysis. The 1955 Allerton House conference at the University of Illinois was an effort to invigorate the method of content analysis, perceived to be in decline at that time (Diefenbach, in press; Pool, 1959). A multidisciplinary gathering organized by George Gerbner at the University of Pennsylvania in 1966 (documented in Gerbner et al., 1969) was a truly amazing blend of individuals with diverse approaches and common goals. The publication of Ole Holsti's (1969) book, *Content Analysis for the Social Sciences and Humanities,* gave researchers a

book carefully based in scientific method. And it reflects true interdisciplinary variety, as shown by his collection of research questions from the 1940s through the 1960s (p. 15); it encompasses music lyrics, literature, political communication, news coverage, advertising, authorship attribution, psychographics, and psychiatry.

After these advances, content analysts seemed to go their separate ways in the 1970s and 1980s. They tended to settle into disciplinary splits that were productive in the sense that they allowed specializations to be established and new areas of expertise to develop (e.g., methods particularly appropriate to communication; Budd, Thorp, & Donohew, 1967; Krippendorff, 1980; Smith, 1978). In the 1990s, the Internet and the World Wide Web changed that. With CONTENT, the e-mail listserv established in 1994 by Bill Evans at Georgia State University, contact was made.[13] Content analysts needed to learn to speak each others' languages again, and this process is apparent in the discussions that take place on the listserv. Today, CONTENT has more than 1,000 subscribers. Most are graduate students and faculty in university social science programs, but the mailing list also includes computer scientists, software developers, market researchers, political consultants, and public health professionals. Participants reside in the United States, Germany, Sweden, New Zealand, Canada, Great Britain, Japan—anywhere the Internet can reach. Bill Evans also went online with the first content analysis resources Web site in 1996; the reader may link to it from *The Content Analysis Guidebook Online* (see Resource 5).

What the 1955 Allerton House Conference participants discovered was a set of commonalities: Attendees across disciplines were concerned with issues related to using content analysis to make inferences about sources or receivers and with the trend toward going beyond "simply counting the frequency of words or other symbols" (Diefenbach, in press, p. 13). The next chapter revives these themes and attempts to incorporate them into a system of thinking, an "integrative" model of content analysis that demands that the researcher think outside of the box of describing content.

Notes

1. There is another "content analysis" in the discipline of chemistry, where the content is a compound for which chemical composition is analyzed.

2. The following are capsule summaries of the indexes and abstracts used in the creation of Box 2.1:

ABI/INFORM (Global edition) provides abstracts from more than 1,300 business and management publications, including 350+ English-language titles from outside the United States. Complete articles for more than 600 sources are also provided. Topics covered in the abstracts and articles include business conditions, trends, corporate strategies and tactics, management techniques, and competitive and product information. Dates covered range from 1971 to the present.

Anthropological Literature index collects citations of articles and essays on anthropology, archaeology, art history, demography, economics, psychology, and religious studies. It includes English- and European-language articles, two or more pages long, published from the late 19th century to present.

Arts & Humanities Citation Index covers the journal literature of arts and humanities disciplines, including archaeology, architecture, art, Asian studies, classics, dance, folklore, history, language, linguistics, literary reviews, literature, music, philosophy, poetry, radio, television and film, religion, and theater. It currently indexes 1,144 discipline-specific journals plus relevant items from over 6,800 science and social science journals. Information from 1980 forward can be accessed through the index.

ComAbstracts (1979 to present) online database contains abstracts of articles published in the communication discipline. Approximately 50 journals are currently represented, with coverage spanning from 1979 to present (though coverage varies, depending on the journal). The full text of the abstracts may be searched by word, phrase, or author and synonyms for many search terms are automatically included in searches unless explicitly overridden (e.g., a search for "youth" will also return "teen" and "adolescent").

Communication Abstracts (1998 to present) includes references to literature in all areas of communication (mass, interpersonal, and more). It currently covers only the period from 1998 to the present.

Education Abstracts covers English-language periodicals, monographs, and yearbooks about the education discipline. Every article (of one column or more) in 423 periodicals and yearbooks is cited, along with books related to education (published in 1995 or later). Topics covered include contemporary education issues, such as government funding, instructional media, multicultural education, religious education, student counseling, competency based, and information technology. Dates covered are 1983 to present, though information from 1983 to 1993 is only indexed; abstracts begin in 1994.

ERIC, which stands for Educational Resources Information Center, is the world's largest source of education-related literature and information. Abstracts of journal articles and documents (e.g., reports, papers, books) on education research and practice are covered. In addition to the usual search features, ERIC includes a target audience code that identifies who a particular piece was intended for. It covers the period from 1966 to present.

HealthSTAR database focuses on Health Services, Technology, Administration, and Research literature. It contains bibliographic citations (plus abstracts, if available) to journal articles, technical and government reports, meeting papers and abstracts, books, and book chapters. The index covers topics with both a clinical focus (e.g., evaluation of patient outcomes and effectiveness of procedures, programs, etc.) and nonclinical focus (e.g., health policy, health services research). Information from 1975 forward can be accessed through the service.

Historical Abstracts includes references to more than 2,000 journals covering aspects of world history from 1450 to present (excluding the United States and Canada). Included are key historical journals, books, and dissertations, as well as social sciences and humanities journals of interest to history students and researchers. Historical Abstracts thus serves as a resource for researchers in fields such as history, sociology, multicultural studies, psychology, women's studies/gender studies, religion, interdisciplinary studies, anthropology and political science. Dates covered range from 1450 to the present.

Humanities Abstracts database cites articles since 1984 from approximately 468 English-language periodicals (journals and magazines) covering fields such as archaeology and classical studies, art and photography, folklore, history, journalism and communications, language and literature, literary and political criticism, music and performing arts, philosophy, and religion and theology. Some full-text articles are included.

Library Literature index includes references from 234 library and information science periodicals, plus books, book chapters, conference proceedings, state journals, theses, and pamphlets. Topics covered include the following: censorship, public relations, preservation, copyright legislation, automation, cataloging and classification, and electronic searching. The database covers the period from 1984 to the present.

MEDLINE database contains bibliographic citations and author abstracts from the fields of medicine, nursing, dentistry, veterinary medicine, health care, and preclinical sciences. More than 3,900 biomedical journals from the United States and abroad are represented; 88% of citations are in English, and 76% have English-language abstracts. The database comes from the National Library of Medicine (NLM), which indexes each article with NLM's controlled vocabulary, MeSH (Medical Subject Headings). The period covered is from 1966 to the present.

Music Index Online covers the journal literature of classical and popular music since 1979. Indexed topics include musicological or organological subjects and also book and record reviews, first performances, and obituaries. Articles from more than 640 international music periodicals are currently represented and organized within, according to an internal subject heading list.

PAIS International database indexes literature relevant to public affairs (e.g., current issues and actions affecting communities, governments, and people; topics that are or may become the subject of legislation). Subject areas covered include economics, political science, public administration, international trade, international law, international relations, demography, and social problems. PAIS International incorporates references to more than 425,000 journal articles, books, reports, Internet material, and more. Publications from more than 120 countries are indexed, in a variety of languages. Dates covered by the database are 1972 to the present.

Periodical Abstracts is a general index providing access to abstracts from general and academic periodicals covering a variety of disciplines. More than 1,800 periodicals are currently represented. Dates covered range from 1986 to the present.

ProQuest Digital Dissertations database contains citations of more than 1.6 million doctoral dissertations and master's theses published from 1861 (the date of the first accepted U.S. dissertation) to present. The full text of more than one million entries can be obtained in paper and microfilm formats, and the full text of 100,000-plus dissertations can be downloaded from the Web site. Although full-text privileges are restricted to institutional subscribers to the service, free 24-page previews of dissertations and theses (from 1997 forward) are freely available. In addition, citations for post-1979 dissertations include 350-word abstracts, and citations from post-1987 theses include 150-word abstracts. Around 47,000 new dissertations and 12,000 new theses are added to the index each year.

PsycINFO database collects citations from publications deemed psychologically relevant by the American Psychological Association, including journal articles, dissertations, reports, book chapters, and books (all English-language). In addition to psychology literature, the database includes literature relevant to communication, education, medicine, law, nursing, and social work. Dates covered range from 1967 to the present, with Historic PsychINFO going back as far as 1887 for some sources.

Social Sciences Citation Index covers the journal literature of a variety of social science disciplines, including anthropology, communication, history, industrial relations, information science and library science, law, linguistics, philosophy, psychology, psychiatry, political science, public health, social issues, social work, sociology, substance abuse, urban studies, and women's studies. Currently, more than 1,725 journals and 3.15 million articles are indexed, plus relevant items from approximately 3,300 scientific and technical journals. The database allows users to access information from 1980 onward, and author abstracts are available from approximately 60% of cited works. In addition, the index also cites references, allowing users to check the cited works in each article as well as the number of times the article has been cited in other works.

Sociological Abstracts collects research citations from sociology and related social and behavioral science disciplines. It includes information from 2,600 journals plus conference papers, books, and dissertations. The database covers research from 1963 to the present; however, only records after 1974 include in-depth, nonevaluative abstracts.

WorldCat, a general online database, includes references to any type of material cataloged by Online Computer Library Center member libraries (including master's theses). Areas covered include arts and humanities, business and economics, conferences and proceedings, consumer affairs and people, education, engineering and technology, general and reference, general science, life sciences, medical and health sciences, news and current events, public affairs and law, and social sciences. Around 40 million records are currently indexed, dating from the 12th century to present.

3. In Box 2.1, years are left blank for which no search is possible for an index. A zero indicates that the index is operable for that year, but no citations to content analysis or text analysis are found. For example, Humanities Abstracts begins with the year 1984; there is one content analysis/text analysis cite for that year.

4. The descriptions were provided by *Harrison's Reports*, a commercial reviewing service available to movie exhibitors. The Reports included a brief plot description for each film released. Dale did a validity check of the Harrison plot summaries.

5. Notice that there was no "why" component to the question. Lasswell's receiver (rather than source) orientation is well noted.

6. Lasswell trained in psychoanalytical methods with Elton Mayo at Harvard, and his subsequent psychoanalytical research was rejected by political scientists and Freudians alike. In the late 1930s, he made plans to establish an institute with Yale anthropologist and linguist Edward Sapir and psychoanalyst Henry Stack Sullivan, a plan cut short by Sapir's death.

7. In Lasswell's seminal work (Lasswell et al., 1949), one of his applications of content analysis was his "propaganda detection," a coding system to be used in support of the McCormack Act, a 1939 law that provided for the registration of foreign agents with the federal government (Spak, 1990). He notes, "In periods of crisis, it is peculiarly necessary to identify enemies of democracy, and to stimulate the members of the public to be on guard in evaluating what is said" (Lasswell et al., 1949, p. 175).

8. Lee and Lee's (1939) seven propaganda devices are (a) name calling, (b) glittering generality, (c) transfer, (d) testimonial, (e) plain folks, (f) card stacking, and (g) band wagon. This classification system is still used today (Severin & Tankard, 1997).

9. Riefenstahl's master works *Triumph of the Will* (1935) and *Olympia* (1937), artistically groundbreaking films documenting the 1934 Nuremberg Nazi Party Rally and the 1936 Berlin Olympics, respectively, are credited by popular and scholarly critics as having stimulated public support for Adolph Hitler (Kracauer, 1947).

10. Sapir (1927) identified five levels for "get[ting] at the personality of an individual" (p. 904) from speech characteristics: (a) voice, (b) voice dynamics (e.g., intonation, rhythm, continuity, speed), (c) pronunciation, (d) vocabulary, and (e) style.

11. Signorielli, Gerbner, and Morgan (1995) provide the following partial list: The Surgeon General's Scientific Advisory Committee on Television and Social Behavior, the National Institute of Mental Health, the White House Office of Telecommunications Policy, the American Medical Association, the Administration on Aging, the National Science Foundation, the W. Alton Jones Foundation, the Screen Actors' Guild, the American Federation of Television and Radio Artists, the National Cable Television Association, the U.S. Commission on Civil Rights, the Turner Broadcasting System, the Office of Substance Abuse Prevention and the Center for Substance Abuse Prevention of the U.S. Public Health Service (p. 279). The Cultural Indicators Project findings are reported in Gerbner (1972), Gerbner and Gross (1976), Gerbner, Gross, Signorielli, Morgan, and Jackson-Beeck (1979), Gerbner et al. (1980), Gerbner, Gross, Morgan, and Signorielli (1986), and Gerbner, Gross, Morgan, and Signorielli (1994).

12. Marshall McLuhan predicted a global village, linked by electronic technologies (McLuhan & Powers, 1989). Fred Williams (1982) refined this prediction to include *multiple* villages defined by communities of interest rather than geography, and this has indeed come to pass.

13. Bill Evans notes that he began CONTENT—the Internet mailing list for news and discussion of content analysis—"because I was lonely. I was at the faculty at Georgia Tech., where I was surrounded by colleagues with interests in research methods and computing. Unfortunately, none of my colleagues had more than a passing acquaintance with content analysis. . . . But I also knew many other researchers at other universities who felt similarly isolated." (William Evans, personal communication, June 2000)

CONTENT provides a forum for discussion among content analysts, and Bill Evans also sends informative content analysis "Publication Alerts" to the listserv subscribers.