1 HISTORY

Empirical inquiries into the meanings of communications date back to theological studies in the late 1600s, when the Church found the printing of nonreligious materials to be a threat to its authority. Such inquiries have since mushroomed, moving into numerous areas and becoming the backbone of communication research. This chapter discusses several stages in the history of content analysis: quantitative studies of the press; propaganda analysis during World War II; social scientific uses of the technique in studies of political symbols, historical documents, anthropological data, and psychotherapeutic exchanges; computer text analysis and the new media; and qualitative challenges to content analysis.

1.1 SOME PRECURSORS

Content analysis entails a systematic reading of a body of texts, images, and symbolic matter, not necessary from an author's or user's perspective. Although the term *content analysis* did not appear in English until 1941 (Waples & Berelson, 1941, p. 2; cited in Berelson & Lazarsfeld, 1948), the systematic analysis of text can be traced back to inquisitorial pursuits by the Church in the 17th century. Religions have always been captivated by the written word, so it is not surprising that the first known dissertations about newspapers were defended in 1690, 1695, and 1699 by individuals pursuing academic degrees in theology. After the advent of the printing press, the Church became worried about the spread of printed matter of a nonreligious nature, and so it dealt with newspaper content in moralizing terms (Groth, 1948, p. 26). Surprisingly, in spite of the rhetorical tradition of ancient Greece, which was normative and oral in orientation, the 17th century contributed very little to the methodology of content analysis.

Probably the first well-documented quantitative analyses of printed matter occurred in 18th-century Sweden. According to Dovring's (1954–1955; see also Krippendorff & Bock, 2009, Chapter 1.1) account, these analyses were undertaken as the result of the publication of the *Songs of Zion,* a collection of 90 hymns of unknown authorship. The collection had passed the Royal Swedish censor, but soon after its publication it was blamed for undermining the orthodox clergy of the Swedish state church. When the collection became popular, it was said to be "contagious" and was accused of aiding a dissenting group. Outstanding in this case is the fact that literary scholars of good reputation participated in the controversy, which crystallized around the question of whether

the songs harbored dangerous ideas and, if so, how. Scholars on one side made a list of the religious symbols in the songs and became alarmed. Those on the other side, however, found the very same symbols in established songbooks and so discounted the claimed difference. Then some scholars noted that the symbols in the songs occurred in different contexts and had acquired meanings that were different from those taught in the official church. A debate arose about whether the meanings should be interpreted literally or metaphorically. The interpretations came to be compared with the results of a German study of the outlawed Moravian Brethren, a religious sect whose members later emigrated to the United States. This process—of revising a method in response to criticism—continued until it became clear to both sides in the debate how the symbols in the *Songs of Zion* differed from the symbols used in the official songbooks and how this (in the end political) phenomenon could be explained. The controversy generated many ideas that are now part of content analysis and stimulated debates about methodology that continue today.

In 1903, Eugen Löbl published in German an elaborate classification scheme for analyzing the "inner structure of content" according to the social functions that newspapers perform. His book, which became well-known in journalistic circles, contributed to the idea of *Publizistik,* or newspaper science, and foreshadowed functionalism, but it did not stimulate empirical investigations.

At the first meeting of the German Sociological Society in 1910, Max Weber (1911; see also Krippendorff & Bock, 2009, Chapter 1.2) proposed a large-scale content analysis of the press, but for a variety of reasons, the research never got off the ground. During the same period, Andrei Markov (1913), who was working on a theory of chains of symbols, published a statistical analysis of a sample of Pushkin's novel in verse, *Eugene Onegin.* These inquiries were discovered only recently or influenced the content analysis literature only indirectly. For example, Weber is celebrated as one of the great sociologists, but his advocacy of the use of content analysis as a method for understanding the mass media is relatively unknown. And Markov's probability theories entered the content analysis literature only through Shannon's mathematical theory of communication (see Shannon & Weaver, 1949), which influenced Osgood's (1959) contingency analysis and cloze procedure.

1.2 QUANTITATIVE NEWSPAPER ANALYSIS

The beginning of the 20th century saw a visible increase in the mass production of newsprint. In the United States, the boom in newspapers created mass markets and interest in public opinion. Journalism schools emerged, leading to demands for ethical standards and for empirical inquiries into the phenomenon of the newspaper. These demands, plus a somewhat simplistic notion of scientific objectivity, were met by what was then called *quantitative newspaper analysis*.

Probably the first quantitative newspaper analysis, published in 1893, asked the rhetorical question, "Do newspapers now give the news?" (Speed, 1893). Its author showed how, between 1881 and 1893, New York newspapers had dropped their coverage of religious, scientific, and literary matters in favor of gossip, sports, and scandals. In a similar but far more simplistic study published in 1910, Mathews

attempted to reveal the overwhelming space that one New York daily newspaper devoted to "demoralizing," "unwholesome," and "trivial" matters as opposed to "worthwhile" news items. By simply measuring the column inches that newspapers devoted to particular subject matters, journalists in the early 20th century attempted to reveal "the truth about newspapers" (Street, 1909). Some believed that they had found a way of showing that the profit motive was the cause of "cheap yellow journalism" (Wilcox, 1900); others became convinced that they had established "the influence of newspaper presentations on the growth of crime and other antisocial activity" (Fenton, 1910). At least one concluded that a "quarter century survey of the press content shows demand for facts" (White, 1924).

Quantitative newspaper analysis seemingly provided the needed scientific ground for journalistic arguments. The respect for numbers has a long history, and facts that could be quantified were considered irrefutable. In a footnote, Berelson and Lazarsfeld (1948) quote from a source published more than 200 years ago:

> Perhaps the spirit of the battle over ratification is best reflected in the creed ironically attributed to each of the contending parties by its opponents. The recipe for an Anti-Federalist essay which indicates in a very concise way the class-bias that actuated the opponents of the Constitution, ran in this manner: "wellborn, nine times—Aristocracy, eighteen times—Liberty of the Press, thirteen times repeated—Liberty of Conscience, once—Negro Slavery, once mentioned—Trial by Jury, seven times—Great men, six times repeated—Mr. Wilson, forty times . . .—put them together and dish them up at pleasure. (p. 9; quoted from *New Hampshire Spy,* November 30, 1787)

Quantitative newspaper analysis led to the development of many valuable ideas, however. In 1912, Tenney (see also Krippendorff & Bock, 2009, Chapter 1.4) made a far-reaching proposal for a large-scale and continuous survey of press content to establish a system of bookkeeping of the "social weather" "comparable in accuracy to the statistics of the U.S. Weather Bureau" (p. 896). He demonstrated what he had in mind with an analysis of a few New York newspapers for different ethnic groups, but his proposal exceeded the scope of what was then feasible. Quantitative newspaper analysis culminated in sociologist Malcolm M. Willey's 1926 book *The Country Newspaper.* In this model study, Willey traced the emergence of Connecticut country weeklies, examining circulation figures, changes in subject matter, and the social role these papers acquired in competition with large city dailies.

When other mass media became prominent, researchers extended the approach first used in newspaper analysis—measuring volumes of coverage in various subject matter categories—initially to radio (Albig, 1938) and later to movies and television. Content analysis in subject matter categories continues today and is applied to a wide variety of printed matter, such as textbooks, comic strips, speeches, and print advertising.

1.3 EARLY CONTENT ANALYSIS

The second phase in the intellectual growth of content analysis, which took place in the 1930s and 1940s, involved at least four factors:

- During the period following the 1929 economic crisis, numerous social and political problems emerged in the United States. Many Americans believed that the mass media were at least partially to blame for such problems as yellow journalism, rising crime rates, and the breakdown of cultural values.
- New and increasingly powerful electronic media of communication, first radio and later television, challenged the cultural hegemony of the newspapers. Researchers could not continue to treat these new media as extensions of newspapers, because they differed from the print media in important ways. For example, users of radio and television did not have to be able to read.
- Major political challenges to democracy were linked to the new mass media. For example, the rise of fascism was seen as nourished by the as-yet little-known properties of radio.
- Perhaps most important, this period saw the emergence of the behavioral and social sciences as well as increasing public acceptance of the theoretical propositions and empirical methods of inquiry associated with them.

In the 1930s, sociologists started to make extensive use of survey research and polling. The experience they gained in analyzing public opinion gave rise to the first serious consideration of methodological problems of content analysis, published by Woodward in a 1934 article titled "Quantitative Newspaper Analysis as a Technique of Opinion Research." From writings about public opinion, interest in social stereotypes (Lippmann, 1922) entered the analysis of communications in various forms. Questions of representations were raised, with researchers examining topics such as how Negroes were presented in the Philadelphia press (Simpson, 1934); how U.S. textbooks described wars in which the United States had taken part, compared with textbooks published in countries that were former U.S. enemies (Walworth, 1938); and how nationalism was expressed in children's books published in the United States, Great Britain, and other European countries (Martin, 1936).

One of the most important concepts that emerged in psychology during this time was the concept of "attitude." It added evaluative dimensions to content analysis, such as "pro-con" or "favorable-unfavorable," that had escaped the rough subject matter categories of quantitative newspaper analysis. Attitude measures redefined journalistic standards of fairness and balance and opened the door to the systematic assessment of bias. Among the explicit standards developed, Janis and Fadner's (1943/1965) "coefficient of imbalance" deserves mention. Psychological experiments in rumor transmission led Allport and Faden to study newspaper content from an entirely new perspective. In their 1940 article, "The Psychology of Newspapers: Five Tentative Laws," they attempted to account for the changes that information undergoes as it travels through an institution and finally appears on the printed page.

The interest in political symbols added another feature to the analysis of public messages. McDiarmid (1937), for example, examined 30 U.S. presidential inaugural addresses for symbols of national identity, of historical significance, of government, and of fact and expectations. Most important, Lasswell (1938), viewing public communications within his psychoanalytical theory of politics, classified symbols into such categories as "self" and "others" and forms of "indulgence" and

"deprivation." His symbol analysis led to his "World Attention Survey," in which he compared trends in the frequencies with which prestige newspapers in several countries used national symbols (Lasswell, 1941; see also Krippendorff & Bock, 2009, Chapter 5.3).

Researchers in several disciplines examined the trends in scholarship, as reflected in the topics that representative journals published. Rainoff's (1929) Russian study regarding physics was probably the first of this kind, but the most thorough analyses were conducted in the field of sociology (Becker, 1930, 1932; Shanas, 1945) and later in journalism (Tannenbaum & Greenberg, 1961).

Several factors influenced the transition from quantitative newspaper analysis, which was largely journalism driven, to content analysis:

- Eminent social scientists became involved in these debates and asked new kinds of questions.
- The concepts these social scientists developed were theoretically motivated, operationally defined, and fairly specific, and interest in stereotypes, styles, symbols, values, and propaganda devices began to replace interest in subject matter categories.
- Analysts began to employ new statistical tools borrowed from other disciplines, especially from survey research but also from experimental psychology.
- Content analysis data became part of larger research efforts (e.g., Lazarsfeld, Berelson, & Gaudet, 1948), and so content analysis no longer stood apart from other methods of inquiry.

The first concise presentation of these conceptual and methodological developments under the new umbrella term *content analysis* appeared in a 1948 mimeographed text titled *The Analysis of Communication Content,* authored by Berelson and Lazarsfeld, which was later published as Berelson's *Content Analysis in Communications Research* (1952). This first systematic presentation codified the field for years to come.

1.4 PROPAGANDA ANALYSIS

Berelson described content analysis as the use of mass communications as data for testing scientific hypotheses and for evaluating journalistic practices. Yet the most important and large-scale challenge that content analysis faced came during World War II, when it was employed in efforts to extract information from propaganda. Before the war, researchers analyzed texts in order to identify "propagandists," to point fingers at individuals who were attempting to influence others through devious means. Fears concerning such influence had several origins. Propaganda was used extensively during World War I (Lasswell, 1927), and the years between the two world wars witnessed the effective use of propaganda by antidemocratic demagogues in Europe. In addition, Americans tend to have deep-seated negative attitudes toward religious fanatics, and the lack of knowledge concerning what the extensive use of the

new mass media (radio, film, and television) could do to people raised concerns as well. According to the Institute for Propaganda Analysis (1937), propagandists reveal themselves through their use of tricks such as "name-calling," employing "glittering generalities," "plain folks" identifications, "card stacking," "bandwagon" devices, and so on. Such devices could be identified easily in many religious and political speeches, even in academic lectures, and this approach to propaganda analysis led to a kind of witch hunt for propagandists in the United States. Theories concerning subliminal messages, especially in advertising, raised widespread suspicion as well.

In the 1940s, as U.S. attention became increasingly devoted to the war effort, the identification of propagandists was no longer an issue. Nor were researchers particularly interested in revealing the power of the mass media of communication to mold public opinion; rather, military and political intelligence were needed. In this climate, two centers devoted to propaganda analysis emerged. Harold D. Lasswell and his associates, having written on political symbolism, worked with the Experimental Division for the Study of Wartime Communications at the U.S. Library of Congress, and Hans Speier, who had organized a research project on totalitarian communication at the New School for Social Research in New York, assembled a research team at the Foreign Broadcast Intelligence Service of the U.S. Federal Communications Commission (FCC). The Library of Congress group focused on analyzing newspapers and wire services from abroad and addressed basic issues of sampling, measurement problems, and the reliability and validity of content categories, continuing the tradition of early quantitative analysis of mass communications (Lasswell, Leites, & Associates, 1965).

The FCC group analyzed primarily domestic enemy broadcasts and surrounding conditions to understand and predict events within Nazi Germany and the other Axis countries and to estimate the effects of Allied military actions on the war mood of enemy populations. The pressures of day-to-day reporting left the analysts little time to formalize their methods, and Berelson (1952) thus had little to say about the accomplishments of the FCC group. After the war, however, Alexander L. George worked through the volumes of reports that resulted from these wartime efforts to describe methods that had evolved in the process and to validate the inferences the researchers had made by comparing them with documentary evidence now available from Nazi archives. These efforts resulted in his book *Propaganda Analysis* (1959a; see also Krippendorff & Bock, 2009, Chapter 1.5), which made major contributions to the conceptualization of the aims and processes of content analysis.

The assumptions that propagandists are rational, in the sense that they follow their own propaganda theories in their choice of communications, and that the meanings of propagandists' communications may differ for different people reoriented the FCC analysts from a concept of "content as shared" (Berelson would later say "manifest") to conditions that could explain the motivations of particular communicators and the interests they might serve. The notion of "preparatory propaganda" became an especially useful key for the analysts in their effort to infer the intents of broadcasts with political content. In order to ensure popular support for planned military actions, the Axis leaders had to inform, emotionally arouse, and otherwise prepare their countrymen and -women to accept those actions; the FCC analysts discovered that they could learn a great deal about the enemy's intended actions by recognizing such preparatory efforts in the domestic press and broadcasts.

They were able to predict several major military and political campaigns and to assess Nazi elites' perceptions of their situation, political changes within the Nazi governing group, and shifts in relations among Axis countries. Among the more outstanding predictions that British analysts were able to make was the date of deployment of German V weapons against Great Britain. The analysts monitored the speeches delivered by Nazi propagandist Joseph Goebbels and inferred from the content of those speeches what had interfered with the weapons' production and when. They then used this information to predict the launch date of the weapons, and their prediction was accurate within a few weeks.

Several lessons were learned from these applications of content analysis, including the following:

- Content is not inherent to communications. People typically differ in how they read texts. The intentions of the senders of broadcast messages may have little to do with how audience members hear those messages. Temporal orderings, individuals' needs and expectations, individuals' preferred discourses, and the social situations into which messages enter are all important in explaining what communications come to mean. Interpretations on which all communicators readily agree are rare, and such interpretations are usually relatively insignificant.
- Content analysts must predict or infer phenomena that they cannot observe at the time of their research. The inability to observe phenomena of interest tends to be the primary motivation for using content analysis. Whether the analyzed source has reasons to hide what the analyst desires to know (as in the case of an enemy during wartime or the case of someone needing to impress) or the phenomena of interest are inaccessible in principle (e.g., an individual's attitudes or state of mind, or historical events) or just plain difficult to assess otherwise (such as what certain mass-media audiences could learn from watching TV), the analyst seeks answers to questions that go outside a text. To be sure, the questions that a content analyst seeks to answer are the analyst's questions, and as such, they are potentially at odds with whether others could answer them and how. Quantitative newspaper analysts made inferences without acknowledging their own conceptual contributions to what they thought they found but actually inferred. Content is not the whole issue; rather, the issue is what can be legitimately inferred from available texts.
- In order to interpret given texts or make sense of the messages intercepted or gathered, content analysts need elaborate models of the systems in which those communications occur (or occurred). The propaganda analysts working during World War II constructed such models more or less explicitly. Whereas earlier content analysts had viewed mass-produced messages as inherently meaningful and analyzable unit by unit, the propaganda analysts succeeded only when they viewed the messages they analyzed in the context of the lives of the diverse people presumed to use those messages.
- For analysts seeking specific political information, quantitative indicators are extremely insensitive and shallow. Even where large amounts of

quantitative data are available, as required for statistical analyses, these tend not to lead to the "most obvious" conclusions that political experts would draw from qualitative interpretations of textual data. Qualitative analyses can be systematic, reliable, and valid as well.

Convinced that content analysis does not need to be inferior to unsystematic explorations of communications, numerous writers in the postwar years, such as Kracauer (1947, 1952–1953) and George (1959a), challenged content analysts' simplistic reliance on counting qualitative data. Smythe (1954) called this reliance on counting an "immaturity of science" in which objectivity is confused with quantification. However, the proponents of the quantitative approach largely ignored the criticism. In his 1949 essay "Why Be Quantitative?" Lasswell (1949/1965b) continued to insist on the quantification of symbols as the sole basis of scientific insights. His approach to propaganda analysis produced several working papers but very few tangible results compared with the work of the FCC group of scholars. Today, quantification continues, although perhaps no longer exclusively.

1.5 CONTENT ANALYSIS GENERALIZED

After World War II, and perhaps as the result of the first integrated picture of content analysis provided by Berelson (1952), the use of content analysis spread to numerous disciplines. This is not to say that content analysis emigrated from mass communication. In fact, the very "massiveness" of available communications continued to attract scholars who looked at the mass media from new perspectives. For example, Lasswell (1941) realized his earlier idea of a "world attention survey" in a large-scale study of political symbols in French, German, British, Russian, and U.S. elite press editorials and key policy speeches. He wanted to test the hypothesis that a "world revolution" had been in steady progress for some time (Lasswell, Lerner, & Pool, 1952). Gerbner and his colleagues pursued Gerbner's (1969) proposal to develop "cultural indicators" by analyzing, for almost two decades, one week of fictional television programming per year, mainly to establish "violence profiles" for different networks, to trace trends, and to see how various groups (such as women, children, and the aged) were portrayed on U.S. television (see, e.g., Gerbner, Gross, Signorielli, Morgan, & Jackson-Beeck, 1979).

Psychologists began to use content analysis in four primary areas. The first was the inference of motivational, mental, or personality characteristics through the analysis of verbal records. This application started with Allport's (1942) treatise on the use of personal documents, Baldwin's (1942) application of "personal structure analysis" to cognitive structure, and White's (1947) value studies. These studies legitimated the use of written material, personal documents, and individual accounts of observed phenomena as an addition to the then-dominant experimental methods. A second application was the use of verbal data gathered in the form of answers to open-ended interview questions, focus group conversations, and verbal responses to various tests, including the construction of Thematic Apperception Test (TAT) stories. In the context of TAT stories, content analysis acquired the status of a supplementary technique. As such, it allowed researchers to use data that they could gather without imposing too much structure on subjects and to validate findings they had obtained

through different techniques. Psychological researchers' third application of content analysis concerned processes of communication in which content is an integral part. For example, in his "interaction process analysis" of small group behavior, Bales (1950) used verbal exchanges as data through which to examine group processes. The fourth application took the form of the generalization of measures of meaning over a wide range of situations and cultures (which derived from individualist notions of meaning or content). Osgood (1974a, 1974b) and his students found numerous applications for Osgood, Suci, and Tannenbaum's (1957) semantic differential scales and conducted worldwide comparisons of cultural commonalities and differences.

Anthropologists, who started using content analysis techniques in their studies of myths, folktales, and riddles, have made many contributions to content analysis, including the componential analysis of kinship terminology (Goodenough, 1972). Ethnography emerged in anthropology, and although ethnographers often interact with their informants in ways that content analysts cannot interact with authors or readers, after ethnographers gather their field notes they start to rely heavily on methods that are similar to those that content analysts use.

Historians are naturally inclined to look for systematic ways to analyze historical documents, and they soon embraced content analysis as a suitable technique, especially where data are numerous and statistical accounts seem helpful. Social scientists also recognized the usefulness of educational materials, which had long been the focus of research. Such materials are a rich source of data on processes of reading (Flesch, 1948, 1951) as well as on a society's larger political, attitudinal, and value trends. In addition, literary scholars began to apply the newly available techniques of content analysis to the problem of identifying the authors of unsigned documents.

On one hand, this proliferation of the use of content analysis across disciplines resulted in a loss of focus: Everything seemed to be content analyzable, and every analysis of symbolic phenomena became a content analysis. On the other hand, this trend also broadened the scope of the technique to embrace what may well be the essence of human behavior: talk, conversation, and mediated communication.

In 1955, responding to increasing interest in the subject, the Social Science Research Council's Committee on Linguistics and Psychology sponsored a conference on content analysis. The participants came from such disciplines as psychology, political science, literature, history, anthropology, and linguistics. Their contributions to the conference were published in a volume titled *Trends in Content Analysis,* edited by Ithiel de Sola Pool (1959a). Despite obvious divergence among the contributors in their interests and approaches, Pool (1959a, p. 2) observed, there was considerable and often surprising convergence among them in two areas: They exhibited (a) a shift from analyzing the "content" of communications to drawing inferences about the antecedent conditions of communications and (b) an accompanying shift from measuring volumes of subject matter to counting simple frequencies of symbols and then to relying on contingencies (co-occurrences).

1.6 COMPUTER TEXT ANALYSIS

The late 1950s witnessed considerable interest among researchers in mechanical translation, mechanical abstracting, and information retrieval systems. Computer languages suitable for literal data processing emerged, and scholarly journals started

to devote attention to computer applications in psychology, the humanities, and the social sciences. The large volumes of written documents to be processed in content analysis and the repetitiveness of the coding involved made the computer a natural but also a difficult ally of the content analyst.

The development of software for literal (as opposed to numerical) data processing stimulated new areas of exploration, such as information retrieval, information systems, computational stylistics (Sedelow & Sedelow, 1966), computational linguistics, word processing technology, and computational content analysis. New software also revolutionized tedious literary work, such as indexing and the creation of concordances. Probably the first computer-aided content analysis was reported by Sebeok and Zeps (1958), who made use of simple information retrieval routines to analyze some 4,000 Cheremis folktales. In a Rand Corporation paper titled *Automatic Content Analysis,* Hays (1960) explored the possibility of designing a computer system for analyzing political documents. Unaware of both these developments, Stone and Bales, who were engaged in a study of themes in face-to-face interacting groups, designed and programmed the initial version of the General Inquirer system. This culminated in a groundbreaking book by Stone, Dunphy, Smith, and Ogilvie (1966) in which they presented an advanced version of this system and demonstrated its application in numerous areas, ranging from political science to advertising and from psychotherapy to literary analysis.

The use of computers in content analysis was also stimulated by developments in other fields. Scholars in psychology became interested in simulating human cognition (Abelson, 1963; Schank & Abelson, 1977). Newell and Simon (1963) developed a computer approach to (human) problem solving. Linguistics researchers developed numerous approaches to syntactic analysis and semantic interpretation of linguistic expressions. Researchers in the field of artificial intelligence focused on designing machines that could understand natural language (with very little success).

In 1967, the Annenberg School of Communications (which later became the Annenberg School for Communication) sponsored a major conference on content analysis. Discussions there focused on many areas—the difficulties of recording nonverbal (visual, vocal, and musical) communications, the need for standardized categories, the problems involved in drawing inferences, the roles of theories and analytical constructs, what developments content analysts could expect in the near future—but the subject of the use of computers in content analysis permeated much of the conference. Stone et al.'s (1966) book on the General Inquirer had just been published, and it had created considerable hope among content analysts. The contributions to the 1967 conference are summarized in a 1969 volume edited by Gerbner, Holsti, Krippendorff, Paisley, and Stone, the publication of which coincided with Holsti's (1969) survey of the field.

In 1974, participants in the Workshop on Content Analysis in the Social Sciences, held in Pisa, Italy, saw the development of suitable algorithms for computer content analysis as the only obstacle to better content analyses (Stone, 1975). Since that time, computational approaches have moved in numerous directions. One has been the development of customizable content analysis packages, of which the General Inquirer was the most important precursor. Attempts to apply the General Inquirer system to German texts revealed that software's English-language biases and led to more general versions of General Inquirers, such as TextPack. The basic ingredient of the General Inquirer and TextPack is a dictionary of relevant words. In the 1980s,

Sedelow (1989) proposed the idea of using a thesaurus instead, as a thesaurus might be more accurate than a dictionary in reflecting "society's collective associative memory" (p. 4; see also Sedelow & Sedelow, 1986). In the 1990s, George Miller initiated a major research effort to chart the meanings of words using a computer-traceable network called WordNet (see Miller et al., 1993). In the 1980s, some authors observed that the enthusiasm associated with large systems that had appeared in the 1960s was fading (see Namenwirth & Weber, 1987), but today the development of text analysis software is proliferating, fueled largely by the historically unprecedented volumes of electronic and digital texts available for content analysis. More recently, Diefenbach (2001) reviewed the history of content analysis by focusing on four specific areas: mass communication research, political science, psychology, and literature.

Naturally, many researchers have compared computer-based content analyses with human-based content analyses. For example, Schnurr, Rosenberg, and Ozman (1992, 1993) compared the Thematic Apperception Test (Murray, 1943) with a computer content analysis of open-ended free speech and found the low agreement between the two to be discouraging. Zeldow and McAdams (1993) challenged Schnurr et al.'s conclusion, however. Nacos et al. (1991) compared humans' coding of political news coverage with data from Fan's (1988) computer-coded approach to the same coverage and found satisfactory correlations between the two. Nacos et al. came to the conclusion that content analysts can best use computers in their research by thinking of them as aids, not as replacements for the highly developed human capabilities of reading, transcribing, and translating written matter. As one might expect, today scholars hold many different opinions regarding the future of the use of computer-based content analysis.

Another development that has influenced how content analysts employ computers in their work is the increasingly common use of word processing software, which provides users with such features as spell-checkers, word- or phrase-finding and -replacing operations, and even readability indices. Although not intended for this purpose, ordinary word processing software makes it possible for a researcher to perform basic word counts and KWIC (keyword in context) analyses, albeit laboriously.

Word processing software is inherently interactive; it is driven by the user's reading of the textual material, not fixed. In the absence of computational theories of text interpretation, content analysts have found the symbiosis of the human ability to understand and interpret written documents and the computer's ability to scan large volumes of text systematically and reliably increasingly attractive. In such collaborations, human coders are no longer used as text-level content analysts; rather, they serve as translators of text or sections of text into categories that emerge during reading and then into a data language (that preserves relevant meanings), which enables various computational algorithms (that cannot respond to meanings) to do housekeeping and summarizing chores. This has given rise to a new class of software designed for computer-aided qualitative text analysis, of which NVivo and ATLAS.ti are two examples. Such interactive-hermeneutic text analysis software is becoming increasingly accessible, especially to students.

The most important stimulus in the development of computational content analysis, however, has been the growing availability of text in digital form. It is very costly to enter handwritten documents, such as transcripts of audio recordings of interviews, focus group protocols, minutes of business meetings, and political

speeches, into a computer. Scanners have vastly improved in recent years, but they are still too unreliable to be used without additional manual editing. In the 1970s, data consortia emerged through which social scientists could share costly data, but the operations of these consortia were marred by a lack of standards and the usually highly specialized nature of the data. Then, in 1977, DeWeese proposed and took the remarkable step of bypassing the costly transcription process by feeding the typesetting tapes of a Detroit newspaper directly into a computer to conduct an analysis of the paper's content the day after it was published. Since that time, word processing software has come to be an integral part of the internal operations of virtually all social organizations; personnel create texts digitally before they appear on paper, use electronic mail systems, and surf the Internet to download materials relevant to their work.

Today, a fantastic amount of raw textual data is being generated daily in digital form, representing almost every topic of interest to social scientists. Electronic full-text databases, to which all major U.S. newspapers, many social science and legal journals, and many corporations contribute all of the materials they publish, are growing exponentially and have become easily available and inexpensive to use online. Add to this the volume of electronic publications, the research potential of the Internet, data available from online multiuser discussions (MUDs) and newsgroups, and online survey systems, which may well replace focus groups and interviews in certain empirical domains, and it is clear that the landscape of how society presents itself has been altered drastically. With more and more people interested in this wealth of digital data, there is a corresponding demand for increasingly powerful search engines, suitable computational tools, text base managing software, encryption systems, devices for monitoring electronic data flows, and translation software, all of which will eventually benefit the development of computer-aided content analysis. The current culture of computation is moving content analysis into a promising future.

1.7 QUALITATIVE APPROACHES

Perhaps in response to the now dated "quantitative newspaper analysis" of more than a century ago or as a form of compensation for the sometimes shallow results reported by the content analysts of 60 years ago, a variety of research approaches have begun to emerge that call themselves *qualitative.* I question the validity and usefulness of the distinction between quantitative and qualitative content analyses. Ultimately, all reading of texts is qualitative, even when certain characteristics of a text are later converted into numbers. The fact that computers process great volumes of text in a very short time and represent these volumes in ways someone can understand does not remove the qualitative nature of the texts being analyzed and the algorithms used to process them: On the most basic level, computers recognize zeros and ones and change them as instructed, proceeding one step at a time. Nevertheless, proponents of qualitative approaches to content analysis offer alternative protocols for exploring texts systematically.

Discourse analysis is one such approach. Generally, *discourse* is defined as text above the level of sentences. Discourse analysts tend to focus on how particular phenomena are represented. For example, Van Dijk (1991) studied manifestations of racism