3 USES AND INFERENCES

This chapter reviews the breadth of applications of content analysis, not in terms of subject matter or the disciplines that engage the technique but in terms of the kinds of inferences that enable content analysts to accomplish their tasks. Several types of logic capable of relating data to their contexts are distinguished here: systems, standards, indices, representations, conversations, and institutions. These outline frameworks for theories that content analysts may draw upon or need to develop.

3.1 TRADITIONAL OVERVIEWS

Content analysis has an important place in the wide range of investigative tools available to researchers. As noted in Chapter 2, content analysis is an unobtrusive technique that allows researchers to analyse relatively unstructured data in view of the meanings, symbolic qualities, and expressive contents they have and of the communicative roles they play in the lives of the data's sources. The combination of these features is unique among research methods. Because virtually all social processes can be seen as transacted through matter that is meaningful to the participants—symbols, messages, images, performances, and organizational phenomena, even nondiscursive practices—the widest use of content analysis is found in the social sciences and humanities, although legal, political, and commercial applications are rising in number as well. Scholars who have surveyed content analysis research have used a variety of categories to describe the growing diversity of research techniques used under the umbrella of content analysis. Janis (1943/1965) offers the following classification:

(1) *Pragmatical content analysis*—procedures which classify signs according to their probable causes or effects (e.g., counting the number of times that something is said which is likely to have the effect of producing favorable attitudes toward Germany in a given audience).

(2) *Semantical content analysis*—procedures which classify signs according to their meanings (e.g., counting the number of times that Germany is referred to, irrespective of the particular words that may be used to make the reference).

   (a) *Designations analysis* provides the frequency with which certain objects (persons, things, groups, or concepts) are referred to, that is, roughly

speaking, subject-matter analysis (e.g., references to German foreign policy).

(b) *Attribution analysis* provides the frequency with which certain characterizations are referred to (e.g., references to dishonesty).

(c) *Assertions analysis* provides the frequency with which certain objects are characterized in a particular way, that is, roughly speaking, thematic analysis (e.g., references to German foreign policy as dishonest).

(3) *Sign-vehicle analysis*—procedures which classify content according to the psychophysical properties of the signs (e.g., counting the number of times the word "Germany" appears). (p. 57)

Leites and Pool (1942; cited in Berelson & Lazarsfeld, 1948) describe four functions of content analysis:

- To confirm what is already believed
- To correct the "optical illusions" of specialists
- To settle disagreements among specialists
- To formulate and test hypotheses about symbols

Berelson (1952) lists 17 uses:

- To describe trends in communication content
- To trace the development of scholarship
- To disclose international differences in communication content
- To compare media or levels of communication
- To audit communication content against objectives
- To construct and apply communication standards
- To aid in technical research operations (to code open-ended questions in survey interviews)
- To expose propaganda techniques
- To measure the readability of communication materials
- To discover stylistic features
- To identify the intentions and other characteristics of the communicators
- To determine the psychological state of persons or groups
- To detect the existence of propaganda (primarily for legal purposes)
- To secure political and military intelligence
- To reflect attitudes, interests, and values (cultural patterns) of population groups
- To reveal the focus of attention
- To describe attitudinal and behavioral responses to communications

Stone, Dunphy, Smith, and Ogilvie (1966) note that although the historical origins of content analysis lie in journalism and mass communication, they found applications of the technique in the following empirical domains:

- Psychiatry
- Psychology
- History
- Anthropology
- Education
- Philology and literary analysis
- Linguistics

I have already mentioned Holsti's (1969) commitment to an encoding/decoding paradigm, which, much like Janis's approach, places message content in the context of communication between senders and receivers. Consequently, Holsti surveys content analyses in terms of three principal purposes:

- To describe *manifest characteristics* of communication—that is, asking what, how, and to whom something is said
- To make inferences as to the *antecedents* of communication—that is, asking *why* something is said
- To make inferences as to the *consequences* of communication—that is, asking *with what effects* something is said

The way in which I categorize the content analyses discussed in this chapter deviates from the ways used by the authors cited above in that it focuses on how researchers *use* content analytic techniques and on how researchers then justify the *inferences* they draw in their analyses. The categories addressed are as follows:

- Extrapolations
- Standards
- Indices and symptoms
- Linguistic re-presentations
- Conversations
- Institutional processes

Not all of the content analysts whose work is reviewed here have explicated the logic of their inferences as we would hope they would. In some cases, this logic is embedded in the notions of meaning that the analysts have subscribed to. In others, the logic can be found in the more or less explicit assumptions that the researchers have made regarding the contexts of their data. Often, this logic remains hidden because researchers take it for granted, presuming that their own tacit knowledge should be obvious to everyone. I have already noted that Berelson

(1952) did not even feel the need to define content. Analysts need to render their assumptions, the logic they employ, examinable. The appropriateness of particular forms of reasoning is an empirical question, of course, not a logical one, and analysts need to judge the appropriateness of their inferences on a case-by-case basis. I have chosen to review content analyses in these terms because such a review will lead most naturally to an understanding of the notion of analytical constructs—but that is another chapter.

3.2 EXTRAPOLATIONS

Extrapolations are inferences of unobserved instances in the intervals between or beyond the observations (data points). Some well-known kinds of extrapolations are interpolations, predictions, extensions, derivations of theorems from other theorems, and systems. Let's take the notion of a system as a general case. A system is a conceptual device, a "complex variable," one might say. The reality that may be described in a system's terms is not part of the definition of a system, although its construction may well be so motivated. Mathematics supplies systems. Its axioms are mathematical, not empirical. Minimally, a system consists of the following:

- *A set of components* whose states are variable

- *Relations* that are manifest in constraints on the co-occurrence of the components' states

- *Transformations* according to which some relations imply other relations in time or in space

One example of a system is our solar system, in which celestial bodies move in relation to each other. The configurations of planets follow a temporal sequence. For someone who knows the system's transformation rules, data on one configuration imply all succeeding configurations. This is a classical Newtonian system. Kinship terminology also constitutes a system, although it is far from deterministic, not as dynamic as the solar system is conceived to be. It defines kin in terms of certain relations between individuals—in English, according to gender, descent, and marriage—and prescribes rights, obligations, modes of address, and so on among kinfolk toward each other. The system allows "extrapolations" in the sense of extending this terminology to individuals entering the system, whether as spouses, children, or adoptees, and it transforms the roles of these individuals relative to each other throughout their lifetimes within the system. Another example of a system is found in the treatment of language as a system of signs, as in the work of Ferdinand de Saussure. The components of language (words and sounds) are thought to be combinable into larger units (sentences and utterances), following grammatical rules. Knowledge of the system enables the knower to generate novel strings of words that are all considered well formed, such as English sentences. Grammar, it should be noted, is not a "natural" kind of system. It is constructed by academic linguists under the assumption that language is a system in its own right.

Some kinds of systems, especially social systems, can be quite complex. The inferences of interest to sociological content analysts are grounded in knowledge

of a society's transformations, which enables analysts to extrapolate features of the system beyond the time and space of available texts—but always within the domain of the system's description. As in the case of grammar, the "rules" by which a social system works are not natural. They are sociological constructions. Although systems can be elaborate, in comparison with other kinds of inferences, extrapolations are relatively simple.

In content analysis, the idea of studying systems goes back to Tenney (1912), who asked,

> Why should not society study its own methods of producing its
> various varieties of thinking by establishing [a] . . . careful system of
> bookkeeping? . . . What is needed . . . is the continuous analysis of a large
> number of journals. . . . The records in themselves would constitute a
> series of observations of the "social weather," comparable in accuracy to the
> statistics of the United States Weather Bureau. (p. 896)

Tenney described systematic relations between subject matter categories within the newspapers he analyzed, noted changes in their distribution over time, and explored, especially, the ethnic characteristics of these publications. He equated the dynamics of press coverage in an entire country with the thinking processes of that country's population, but he lacked methods that were adequate to process the great volume of data that the construction of such a system would require. Tenney made his proposal not only well before computers existed but also before systems theory had been developed.

Rapoport (1969) prepared the ground for a systems theory of "verbal corpuses" as he sought answers to questions such as what it means to describe a large body of verbal data as behaving, changing, and evolving, and what the suitable components, relationships, and laws of interaction within such corpuses might be. Although he was aware that our symbolic world both mirrors and constitutes human existence, and that it can be both enriched in talk and polluted by institutional policies, he suggested that researchers could most fruitfully pursue the large-scale study of verbal corpuses, at least to start out with. This study would be done without reference to speakers, symbol users, and meanings—that is, as a system with its own autonomous regularities. From this perspective, content analysis could be said to probe selectively into what Vernadsky (1945) called a "noosphere," a sphere of human knowledge distinct from the "biosphere" in which humans live qua organisms.

3.2.1 Trends

The prototype of a systems approach in content analysis is the extrapolation of trends. In one of the earliest content analyses ever conducted, Speed (1893) compared several New York dailies published in 1881 with the same newspapers published 12 years later and observed changes in the frequencies of subject matter categories. Of course, data on only two points in time hardly lend themselves to solid predictions, but Speed's lamenting of the continuing decline of newspaper coverage of literary matters and the increase in gossip, sports, and fiction, raising the question of where this trend would lead, is a clear indication of his desire to predict how newspaper publishing

was changing. Lasswell (1941) proposed a study and presented preliminary findings on trends in the frequencies with which references to various countries occurred in different national presses. Loeventhal (1944) studied the changing definition of heroes in popular magazines and found a drift (still ongoing today) away from working professionals and businessmen as heroes and toward entertainers.

Other trend studies have concerned values in inspirational literature, advertising themes, and political slogans, as well as the frequency of the use of the word *mainstreaming* in several subcategories of educational research (Miller, Fullmer, & Walls, 1996). Researchers have also undertaken numerous analyses of trends in scholarly literature, from sociology (Shanas, 1945) to content analysis (Barcus, 1959), to ascertain the directions in which particular fields seem to be moving. Shanas (1945), for example, analyzed emerging interests in the field of sociology in the United States by examining the distribution of articles in the *American Journal of Sociology* over a 50-year period. Scholars in many academic disciplines have, from time to time, made efforts to review their literature to assess the directions in which their fields were moving and to identify new frontiers. One of the more extensive content analyses conducted to date using a time-series analysis approach is Namenwirth's (1973) analysis of value changes in U.S. political party platforms over a 120-year period. Namenwirth and Weber (1987) also applied time-series analysis to a study of all speeches made by British monarchs between 1689 and 1972 (see also Krippendorff & Bock, 2009, Chapter 3.8). Both studies revealed two independent cycles of value changes in the data, a short-term cycle and a long-term cycle (see Figure 10.5). Such findings amount to descriptions of the dynamics of autonomous systems. Thome and Rahlf (1996) analyzed these same data using a "filtering" methodology instead of time-series analysis, but both methods operate within the notion of a system that enables the analyst to interpolate between data points and extrapolate its ups and downs into the future.

Political party platforms, policy positions, and campaign materials are natural candidates for this kind of content analysis because they are recurrent and there is an interest in knowing what comes next. The European Consortium for Political Research, constituted in 1979, has undertaken numerous trend studies of how different political systems behave over time, in various dimensions, and concerning a variety of emerging issues (see, e.g., Budge, Robertson, & Hearl, 1987). For more than two decades, consortium researchers have coded nearly 2,000 party manifestos using a single coding scheme based on 56 categories and performed numerous trend analyses; they are now experimenting with computerizing this approach (Pennings & Keman, 2002).

3.2.2 Patterns

Another kind of content analysis involves the extrapolative use of patterns. In folklore, for example, researchers have conducted structural analyses of riddles, proverbs, folktales, and narratives with the aim of identifying patterns that have a high degree of commonality within genres, regardless of particular contents (Armstrong, 1959), and can therefore be regarded as generative of those genres. Such analysts begin by identifying the constituent elements within a body of literature and then seek to describe the logic that relates these elements. Thus Sebeok and Orzack (1953), analyzing Cheremis charms, found that in such charms, a

"purely factual statement" about the world is followed by a "motif of an extremely improbable eventuality." Labov (1972) found a set of components that accounted for the narratives he had elicited and considered these the building blocks for the construction of narratives generally.

Another example of the extrapolation of patterns is the analysis of genealogies within a body of literature through citation networks. Scholarly works tend to cite previous publications on which they relied or which they criticized, which in turn cite earlier publications, and so on. Citation networks enable the analyst to trace present ideas through these networks to their origins, and what is equally interesting, to infer the importance of scholars or particular conceptions by the number of works that cite them or the networks they grow. Citation networks enable content analysts to identify separate discourse communities who generate tightly interconnected literatures without much connection to the literature of other discourse communities. Or they identify canonical texts that everyone in a particular field has read and draws upon. Garfield (1979) relied on this simple idea when he developed his "citation index" as an alternative to information retrieval by keywords. Communication researchers have charted communication channels among members of organizations as senders and receivers and have analyzed those connections in terms of typical network features that organizations tend to reproduce. On a totally different level, word co-occurrences within sentences or paragraphs have also revealed networks of association that individual contributors of a discipline mindlessly reproduce, patterns that can permeate a genre.

Combined interest in trends and patterns has led to many interesting content analyses. Bales's (1950) "interaction process analysis" yielded patterns of communication, evaluation, control, decision making, tension reduction, and reintegration, all of which were identified within 12 basic categories of verbal exchanges in small groups. Holsti, Brody, and North (1965) studied public statements made by major decision makers in the United States and the Soviet Union during the 1962 Cuban missile crisis and distinguished perceptions and expressions within these successive statements that they described in terms of Osgood's semantic differential dimensions: evaluative, strength, and potency. With the help of a dynamic interdependency model, Holsti et al. found that these data proved moderately predictive of the pattern of emotional responses each group of decision makers made to the other.

Possibly the most fruitful recent reliance on patterns has emerged in studying networked text in social media and on the Internet. Obviously, but not too often taken seriously, everything said is said to someone else and establishes a relationship. When Chapter 2 suggested that content analysts treat texts neither (or at least not exclusively) in terms of their authors' intentions, as many literary scholars do, nor as symbolic representations of physical phenomena, as many semioticians do, but in the contexts of their use, this implies a conception of texts that circulate in communities, enliven interconnected realities, and do something that content analysts could possibly infer. Social media can be considered micro blogs in which users create or reproduce texts and disseminate them within networks of friends or communities that form around particular texts—of identity and of coordinating actions, like the Arab spring, which is known to have resulted from the high-speed, emotionally engaging, and politically focused communication on the Internet. (See Section 11.3.4 for the use of software to reveal networked texts enabled by hyperlinked URLs, Twitter use, and what can be inferred from them.)

3.2.3 Differences

Differences are central to all communication studies. Gregory Bateson (2000, p. 459) defined "the elementary unit of information (as) a difference which makes a difference" and demonstrated the explanatory power of this human-centered conception of information by tracing the flow of differences through social systems of actions. The differences of interest here also include differences between communicators; the communication of ethnic, gender, religious, or national differences; the polarization of politics; and their consequences. For example, content analysts have studied how unlike expectations of what an audience is composed of changes how politicians address them. In one of the earliest content analyses, Berelson and Salter (1946) compared the racial composition of the population of fictional characters in magazines with that of the U.S. population. This comparison may be too simplistic. If the mass media would be asked to accurately represent the population of the public, we would not be able to see magicians, learn about breakthroughs by outstanding scientists, and hear exceptional entertainers. However, differences in the news coverage of public issues do have consequences on differences in the public domain. They have been correlated with editorial endorsements (Klein & Maccoby, 1954) and in covering civil rights issues with newspapers' geographic location, ownership, and political orientation (Broom & Reece, 1955). Differences in newspaper content have been correlated with whether or not newspapers face competition within their regions (Nixon & Jones, 1956).

Gerbner (1964) demonstrated how different ideological and class orientations are reproduced in the stream of French news media messages in the reporting of an apolitical crime. Researchers have also shown how the sources of political messages become aligned with the audience expectations, by comparing, for example, the political speeches that John Foster Dulles made before different kinds of groups (Cohen, 1957; Holsti, 1962). Research has linked differences in television sports reporting of men's and women's athletics to prevailing cultural values (Tuggle, 1997) and revealed differences in works of fiction written for upper-, middle-, and lower-class readers (Albrecht, 1956) as well as in advertisements in magazines with predominantly Black and predominantly White readerships (Berkman, 1963). Allport and Faden (1940) examined how the published newspaper content related to the number of sources of information available to newspapers. How stories changed as they were translated from one medium to another was illustrated in Asheim's (1950) analysis of what happens when books are adapted to become movie scripts. Similar are studies that have compared scientific breakthroughs, discovering how unanticipated differences appear in popular media.

The Hoover Institution's study titled Revolution and the Development of International Relations (RADIR) combined the analysis of differences between media and the analysis of trends. The RADIR researchers identified so-called key symbols such as *democracy, equality, rights,* and *freedom* in 19,553 editorials that appeared in American, British, French, German, and Russian prestige newspapers during the period 1890–1949. Analyses of these data led Pool (1951) to correlations that he sought to generalize. He observed, for example, that proletarian doctrines replaced liberal traditions, increasing threats of war are correlated with growths in militarism and nationalism, and hostility toward other nations is related to perceived insecurity. Although these symbols refer to aspects of political realities and never

left their connections to the contexts in which they were published and read, Pool and other researchers did not need these references to conduct their analyses. They tried to establish which differences were maintained over time, which differences increased or decreased whether correlated or not, and how they compensated for or amplified each other. For example, Pool (1952b) observed that symbols of democracy become less frequent where representative forms of government are accepted rather than in dispute. It should be noted that the knowledge of whether a particular form of government is generally accepted or in dispute comes from outside the system of selected symbols the RADIR researchers were studying. To the extent external variables can explain a system's behavior, in the form of the contributing conditions illustrated in Figure 2.2, such systems are not entirely autonomous. However, content analysts who study such systems have some analytical freedom to include categories of dissent, defiance, and struggle in their analysis, even if they are not recognized within such systems as long as they explain how a system copes with such differences.

In a very different approach, Gerbner and his colleagues accumulated a very large database on television violence in fictional programming that enabled them to make extrapolations (recommendations) of interest to policy makers (see, e.g., Gerbner, Gross, Morgan, & Signorielli, 1994; Gerbner, Gross, Signorielli, Morgan, & Jackson-Beeck, 1979). Gerbner's (1969) "message systems analysis" proposes to trace the movement of mass-media culture through time by means of a system consisting of four kinds of measures of any category of content (component):

- The frequencies with which a system's components occur, or "what is"

- The order of priorities assigned to those components, or "what is important"

- The affective qualities associated with the components, or "what is right"

- The proximal or logical associations between particular components, or "what is related to what"

One might question Gerbner's equation of frequencies with "what is," how stable these quantitative measures really are, and whether the system is sufficiently autonomous. The point, however, is that any system of measurements, when observed long enough, will allow analysts to make predictions in the system's own terms, whatever they mean.

Simonton (1994) has made an interesting and rather unusual use of the content analysis of systems of differences in his analysis of musical transitions in melodies. He analyzed 15,618 melodic themes in the works of 479 classical composers working in different time periods. Simonton was interested in the relationship between originality and success, and he inferred originality from the unusualness of the transitions in particular works and for particular composers relative to the pool of all melodic themes. For example, he found that Haydn's "Symphony no. 94" employs transitions found in 4% of the theme inventory, whereas Mozart's "Introduction to the Dissonant Quartet" uses transitions that occur in less than 1% of this inventory.

Up until the civil rights movement in the United States gained momentum in the 1950s, its Afro-American population was almost invisible on television. Berelson and Salter (1946) showed that much. In his dissertation, Cedric Clark (1966) observed that Afro-American presence in the public media had increased in numbers but

merely as negative contrast to White leads, as the target of demeaning comments and ridiculing jokes. Public dissemination of such differences, particularly to audiences who do not have direct experiences with minorities, creates expectations of differences that, when enacted, reproduce the very differences displayed in the media. Fictional differences tend to become real in the form of self-fulfilling prophesies. This process has been exploited by totalitarian regimes to create internal or external enemies they need to fight in order to maintain the fiction of their own superiority. It can be found also in contexts where ethnic, religious, national, and gender prejudices are created and amplified.

A well-received study of intelligence in American society provides a good example of ignoring the effects of publishing differences. Herrnstein and Murray (1994) applied the standard IQ test to a cross section of the U.S. population, partitioned by race, socioeconomic status, job performance, criminal records, family stability, and more. They found that its African-American population was less intelligent, less achieving economic success, more prone to commit crimes, and so on. Moreover, they observed little change in IQ from one generation to the next, which led them to conclude that intelligence is genetically fixed. The two authors did not reflect on where they learned to make the distinctions they were applying to their data. They made no references to how White-dominated mass media in the United States presented members of the Black community. They did not acknowledge that IQ tests embodied the values of the educated White population and ignored proposals of other conceptions of intelligence, but most importantly, that their findings could have been explained by how this minority is treated by a majority enacting their expectations in hiring practices, admissions to higher education, and suspecting Blacks to be prone to commit criminal offenses. Widely communicating such differences and enacting them amounts to a circular causality that makes it hard for the population of African-Americans to escape that self-validating confinement. Herrenstein and Murray's *Bell Curve*, claiming genetic explanations of population differences when sociolinguistic ones are at their core, merely feeds these prejudices (Naureckas, 1995). Critical content analysts ought to be aware of the contexts in which their analyzed texts reside, and this awareness ought to include the effects of their own publications.

Although studying the textual differences in the context of their causes and consequences is a fruitful application of content analysis, it is not always easy to find appropriate texts to support such inferences. For example, Silicon Valley and mathematics-intense fields like engineering and computer science have been accused of thriving in a sexist culture, without much proof. Representatives of these disciplines blame lack of mathematical competence or disinterest among women for this situation. In economics, too, gender inequality is well documented (Wolfers, 2017) but not what sustains it. What economists actual say when hiring or promoting women candidates to faculty positions takes place largely behind closed doors. Alice H. Wu (2017) took advantage of the intersection of two technologies, online messaging and computer-aided content analysis, and scraped over 1,143,416 posts across 131,913 threads between 2014 and 2016 from Economics Job Market Rumors (EJMR), an online forum where economists in the job marked post news, rumors, and advice about open academic positions. Posts are anonymous, which removes the constraints of political correctness, but makes it difficult to identify the gender of their

contributors. Wu therefore focused on what posts were about. She used four levels of gender-identifiers, starting with personal pronouns "he" and "she" and ending with "guy" and "bro" among seven others. About 31% of the posts were female related, and 61% concerned males. She used a machine learning algorithm (see Section 11.4.2) to identify the 30 most prominent words that differentiated posts discussing females and males. A comparison of these two lists is shocking. Whereas the words that are unique to female posts refer to body parts and unspeakable sexist descriptions, the words that are unique to male posts refer to economic concepts, professional designations, and publications. She also grouped the 10,000 most frequent topic words into 15 categories, among them academic/professional (economics, academic, and professional) and personal/physical (marital status, attractiveness, and gender related attributes), and found additional evidence for the above gender inequality. Posts about females contained 43% fewer academic or professional terms than expected and 192% more personal terms and physical attributes. Among the most prominent authors in economics, men outnumber women by 12 to 1. Several women economists who read her work reported experiences that resonated with her findings (Wolfers, 2017).

Textual differences matter. Their analyses have to address the richly interactive textual dynamics and the social realities they bring about (Krippendorff, 2005).

3.3 STANDARDS

Humans measure observed phenomena against standards to establish (a) the kinds of phenomena they are (identifications), (b) how good or bad the phenomena are (evaluations), and (c) how close the phenomena come to expectations (judgments). I discuss each of these three uses of standards below. The facts that identities do not reveal themselves (they require someone identify them as such), that evaluations are not objective or natural (they are the products of someone's values), and that audits by themselves are inconsequential (unless someone can invoke institutional consequences) reveal that standards facilitate inferences of a certain kind. In content analysis, standards are often implicit. People are quick to be for or against something without any clear idea of why. As suggested in Chapter 2, content analysts should take care to make explicit why they infer what they do, and this includes defining the standards they apply in their studies.

3.3.1 Identifications

Identification concerns *what something is,* what it is to be called, or to what class it belongs. Identifications are "either/or" inferences—that is, something either is or is not of a certain kind. Most basically, all computer text analyses start with the identification of character strings, not meanings. Any two strings are either the same as or different from each other. In his above-cited typology of content analyses, Janis (1943/1965) calls one type "sign-vehicle analysis." In this type of analysis, researchers use procedures that classify content according to the psychophysical properties of the signs (e.g., by identifying the word *Germany* and then counting how often it appears). Dibble (1963), who analyzes the kinds of inferences that historians habitually make in their work, includes "documents as direct indicators" as one kind of inference. For

example, suppose a historian wants to know whether the British ambassador to Berlin communicated with England's foreign ministry the day before World War I began; a letter from the ambassador in the file of that ministry would provide direct evidence of its having been sent and received. Because identifications are often obvious, it is easy to overlook their inferential nature. In content analysis, the simplest task requires that a decision be made concerning whether something has occurred, was said, or has been printed. For example, when officials of the Federal Communications Commission are alerted that certain four-letter words have been broadcast over the public airwaves, they need definite proof that those words have been aired before they can consider suspending the offending station's broadcasting license. Identifications are rarely so simple, however.

The legal system's use of content analysis as an evidentiary technique provides us with many examples of identifications (see "Content Analysis," 1948; Lashner, 1990). Tests aimed at establishing whether a particular publication is defamatory, whether a given political advertisement is based on facts, whether a certain signature is real, and whether a given painting is the work of a particular artist all involve either/ or-type inferences about identities or class memberships, but not all of them are simple and obvious. For example, to identify a statement as defamatory in the context of a legal proceeding, an analyst must show that all components of the applicable legal definition of defamation are satisfied.

3.3.2 Evaluations

Well before the term *content analysis* appeared, at a time when media research was equated with the journalism-inspired premises of quantitative newspaper analysis, the evaluation of press performance was an important issue, as it still is. Early concerns about changes in newspaper publishing (Speed, 1893), which surfaced in public criticisms of increases in the coverage of "trivial, demoralizing, and unwholesome" subject matter at the expense of "worthwhile" information (Mathews, 1910), were certainly motivated by largely unquestioned ideals, evaluative standards, and norms couched in seemingly objective frequency measures. Some cultural critics today may share the concerns expressed by the authors of these early studies, but journalism has changed in the intervening years and has shown itself to be responsive to the evolving cultural climate and to shifting political and economic conditions.

Evaluative studies of newspaper reporting have focused largely on two kinds of bias: the bias in accuracy (truth) of reporting and the bias in favoring one side of a controversy over the other. For example, Ash (1948) attempted to determine whether the U.S. public was given a fair opportunity to learn about both sides of the controversy that accompanied the passage of the Taft-Hartley Labor Act. Accuracy in reporting and favoritism in reporting can be difficult to separate, however. During election campaigns, for instance, most politicians allege that some segments of the media display bias in their election coverage. The more popular candidates, who enjoy frequent attention from the press, tend to complain about inaccuracies in reporting, whereas the less popular candidates, struggling for publicity, are more likely to complain about inattention. Because journalists are committed to being fair to all sides in their reporting, many are defensive when the press is accused of taking sides and take the measurement of bias quite seriously.

In practice, evaluative studies of journalistic practices have not solved the now century-old problem of the lack of unquestionable criteria. Janis and Fadner

(1943/1965) sought to put this deficiency to rest with their publication of a coefficient of imbalance, in which

f = the number of favorable units,

u = the number of unfavorable units,

r = the number of relevant units = $f + u$ + the number of neutral units, and

t = the total number of units = r + the number of irrelevant units.

The "coefficient of imbalance C," which Janis and Fadner derived from 10 propositions intended to capture the prevailing intuitions regarding (im)balance in reporting, measures the degree to which favorable statements, f, outnumber unfavorable statements, u, relative to the two ways of assessing the volume of a text, r and f:

$$C = \begin{cases} \dfrac{f^2 - fu}{rt} \text{ when } f \geq u \\ \dfrac{fu - u^2}{rt} \text{ when } f < u \end{cases}$$

This coefficient ranges in value from –1 to +1. It is a good example of an evaluative standard that enables the kind of inferences we often make without much thinking: It defines an ideal (here a balance between positive and negative evaluations), and it measures deviations from that ideal in degrees (here in either the positive or the negative direction). The reality of evaluative standards is far from clear, however. For example, whether journalists can always be impartial is an unsettled issue; some would argue that there are circumstances under which they may not have to be or under which impartiality may not be possible. In the last days of Nixon's presidency, for example, it was difficult for journalists not to take the side of the public. And so it is in situations of war, where loyalty tends to outweigh fairness to both sides. To give one's nation's enemies a fair hearing might be an intellectual challenge, but in practice, it is utterly unpopular. In the early 1960s, Merrill (1962) tried to differentiate dimensions of evaluative standards for journalistic practices. He proposed a battery of evaluative criteria to be applied to journalistic presentations (attribution bias, adjective bias, adverbial bias, contextual bias, photographic bias, and outright opinion), but his catalog is far from complete.

To assess accuracy in reporting, one must have standards against which to judge representations. Insofar as the reality we know is always already described, accuracy amounts to correspondence with sources that are deemed authentic. In a landmark study, Berelson and Salter (1946) compared the racial composition of the population of fictional characters in magazines with that of the U.S. population. The statistical operationalization of "representativeness" that they used has also been employed in many subsequent evaluative studies (Berkman, 1963). But whether the population of fictional characters in magazines, in plays, or in television programs should be statistically representative of the audience in characteristics such as ethnicity, age, occupation, and artistic capability remains debatable. The community television projects of the 1970s died precisely because audience members did not find it particularly entertaining to look into the lives of their ordinary neighbors. The

"reality" TV shows of today may give the impression of being representative of real life, but they actually amount to contrived games played by carefully selected people. A bit less controversial are comparisons of the contents of narratives with those of other narratives. The Council on Interracial Books for Children (1977) has proposed and demonstrated a method for evaluating history texts in the United States by comparing the information in them with known historical facts. Here too, however, matters are not as simple as they seem. In the presentation of history, some selectivity is unavoidable; such evaluative efforts should aim to discover systematic exclusions and overstatements, not variations around an ultimately arbitrary standard. In journalism, the standard of truthful reporting is almost universally subscribed to, but it often conflicts with journalists' responsibility for the consequences of their reporting—for example, preventing fair trials, stimulating public fears, hyping people into action, and creating scandals.

3.3.3 Judgments

Like identifications and evaluations, judgments are based on standards but with the additional provision that they are *prescribed or legitimated by institutions,* and research using such standards tends to have institutional implications. For example, when the United States Federal Communications Commission (FCC) grants licenses to television stations, the stations are obligated to maintain certain proportions of news, community, and public service programming; that is, the FCC sets explicit criteria with which broadcasters must comply. Content analysts have measured the proportions of different kinds of programming aired on some stations and, in effect, have influenced FCC decisions regarding the status of the stations' broadcasting licenses.

Social scientists have long been fascinated with social deviance, and many have theorized about crime, pornography, obscenity, and the like. In doing so, they have influenced the community of their peers and undoubtedly affected public opinion. However, for content analyses to have institutional implications, their results must be presented in the target institutions' terms; otherwise, they do not have any effect. Content analysts may study such social problems as plagiarism, discriminatory communication practices, and the effects of fictional programming on particular kinds of crimes, but their findings are not likely to support judgments with consequences unless the researchers use the concepts, categories, and language of laws, enforceable agreements, or other institutional standards that are applicable to the institutions concerned with these problems. For example, organizational communication researchers are often asked to perform so-called communication audits of industrial or business organizations, in which they ask what is being said, how, and to whom, and what function it serves. Such an audit is usually driven not by scientific curiosity or public concerns but by expectations from within the organization that the results will be useful, solve problems, or inform effective actions. Early communication audits often failed because they were conducted by academics who measured their findings against communication theories that had little to do with how organizations have to function. If the results of organizational communication research are to lead to consequences, they must be couched in the studied organization's terms and be measured against the standard of communication structures known to be successful.

3.4 INDICES AND SYMPTOMS

An index is a variable whose significance rests on its correlation with other phenomena. According to the semiotician C. S. Peirce, an index must be causally connected to the event it signifies, as smoke indicates fire. This presumes an underlying mechanism such that the relation between an index and what it signifies is a matter of necessity rather than convention (symbol) or similarity (icon). Indices are so conceived in medicine, where they are called *symptoms.* To diagnose, a physician looks for visible or measurable manifestations of an illness. However, even in medicine, symptoms have their histories, and medical practitioners must be educated to recognize them for what they are, which makes symptoms a property of the institution of medicine as much as of the phenomena the symptoms are supposed to indicate. In the social domain, where physical mechanisms (causalities) tend to be absent, the observer-dependent nature of indices is even more prominent. As Rapoport (1969) has noted, "An index . . . does not depend on (or should not be confused with) the physical entities or events from which it is derived" (p. 21).

In content analysis, indices of unobservable or only indirectly accessible phenomena are most common. Typically, analysts use measures of textual (verbal and paralinguistic), visual (gestural and pictorial), and communicational characteristics to address extratextual phenomena. For example, the ratio of disturbed speech to normal speech (speech-disturbance ratio) may serve as an index of a patient's anxiety during psychiatric interviews (Mahl, 1959); the frequency of a category of assertions or images related to action, goals, and progress is understood to indicate their producer's achievement motive (McClelland, 1958); and the frequencies of expressed concerns for an issue and the typographical positions of its expressions in a medium (e.g., in newspapers: size of headlines, front or inside pages, lead paragraphs of stories or mere mentions) are seen as indices of the amount of public attention to that issue (e.g., Budd, 1964). Gerbner et al. (1979) created a television violence index based on the numbers of violent scenes in fictional TV programs. Krendel (1970) developed an index of citizen dissatisfaction based on letters of complaint to city halls. Flesch's (1948, 1951, 1974) "readability yardstick" is derived through a formula that, after several incarnations, responds to two factors: average sentence length (in number of words) and average number of syllables per word. Danielson, Lasorsa, and Im (1992) used Flesch's criteria in their comparison of the readability of newspapers and novels. Government contractors are required to apply a version of Flesch's yardstick before finalizing instructions to military personnel, and insurance companies use it to evaluate contracts. Hawk (1997) extended Flesch's criteria for readability to evaluate the "listenability" of television news. Jamieson (1998) has constructed a campaign conduct index that takes into account Americans' expressed concerns about how much money politicians spend on campaigns, what candidates say to get elected, candidates' ethics and morals, and the proportion of negative ads used in political campaigns. Broder's (1940) adjective-verb ratio has been employed as an index of schizophrenia (Mann, 1944), and above-chance co-occurrences of nouns have been interpreted as indicators of associations in speakers' and receivers' minds (Osgood, 1959).

In mass communication research, five indices have had a long history of use:

- *The presence or absence* of a reference or concept is taken to indicate the source's *awareness* or *knowledge* of the object referred to or conceptualized.

- The *frequency* with which a symbol, idea, reference, or topic occurs in a stream of messages is taken to indicate the *importance of, attention to,* or *emphasis on* that symbol, idea, reference, or topic in the messages.

- The *numbers of favorable* and *unfavorable characteristics* attributed to a symbol, idea, or reference are taken to indicate the *attitudes* held by the writers, the readers, or their common culture toward the object named or indicated.

- The kinds of *qualifications*—adjectives or hedges—used in statements about a symbol, idea, or reference are taken to indicate the *intensity, strength,* or *uncertainty* associated with the *beliefs, convictions,* and *motivations* that the symbol, idea, or reference signifies.

- The *frequency of co-occurrence* of two concepts (excluding those that have grammatical or collocational explanations) is taken to indicate the *strength of associations* between those concepts in the minds of the members of a population of authors, readers, or audiences.

The use of such easily computable quantities as indices is not without its problems. Chomsky (1959) took Skinner to task for suggesting that promptness of response, repetition, and voice volume are natural indices of the intensity of motivation and that meanings can be discerned from the co-occurrence of words with the objects they refer to. He observed that most words are uttered in the absence of what they mean. Rapoport (1969) compares two hypothetical women, each of whom has just received a luxurious bouquet of flowers. The first woman, upon seeing the flowers, shouts, "Beautiful! Beautiful! Beautiful! Beautiful!" at the top of her lungs, thus giving evidence, according to Skinner's criteria, of a strong motivation to produce the response. The second woman says nothing for 10 seconds after she first sees the flowers, then whispers, barely audibly, "Beautiful." Frequency and voice volume would not be good indications of the importance of these flowers or, in Skinner's terms, the motivation to respond.

In content analysis, as in many social scientific inquiries, researchers often simply *declare* indices without demonstrating their empirical validity, especially when the phenomena to be indicated are abstract and far removed from validating data. Obviously, a researcher would not declare a measure to be an index if his or her claim is unlikely to be convincing (i.e., to have face validity) to scientific peers. Simple declarations, however, do not constitute an index as defined above. A declaration is discursive in nature and should not be confused with a correlation between an index and what it claims to indicate. A correlation needs to be demonstrated or at least hypothesized, so that it is testable in principle. Take, for example, a researcher's declaration that the frequency of violence in TV fictional programming is a measure of attention to violence (in real life). To make this claim, the researcher must first clarify whose attention this frequency is supposed to indicate. The authors' or editors'? The audience members actually exposed to the violence so measured or the audiences that producers had in mind attracting, the public at large, or the culture in which these kinds of mass communications are circulating? Given the target of the intended inferences, the researcher must also describe how the attention to be indicated will manifest itself—directly (by observation of TV-related violence) or

indirectly (by correlation with other observable phenomena, such as [in]tolerance for otherness, domestic/disciplinary violence, or crime rate). Counting, emphasizing, paying attention to, and expressing concerns about something are four wholly different things. Their correlation is an empirical question.

Quantification is not an end in itself. Researchers must distinguish between quantifications that lead to the testing of a statistical hypothesis and quantifications that indicate something other than what is counted. These two uses are often confused in the early content analysis literature. For example, in his famous essay "Why Be Quantitative?" Lasswell (1949/1965b) celebrates quantification as the only path to scientific knowledge, by which he means the testing of statistical hypotheses; however, in most of his content analyses, Lasswell used frequency measures as declared indices of extracommunicational phenomena.

In a study of the indicative power of frequencies of mentions, a student of mine used a book on U.S. presidents that was written by a scholar who was available on our university's campus. The student examined the book thoroughly, counting the numbers of mentions of the different presidents; the numbers of chapters, pages, and paragraphs in which each president is mentioned; and the numbers of sentences devoted to each president. He then asked the author to rank the U.S. presidents according to their importance and according to their contributions to U.S. history. He also asked the author how other scholars might rank the presidents and how the public might rank them. Finally, the student even asked the author how much attention he thought he had paid to each of the presidents in his book. Surprisingly, all correlations were very low, to the point that probably none of the measures could serve as a valid index of the author's attention or emphasis. The tentative insight we may derive from this exploratory study is that frequencies may not be good indicators of conceptual variables, such as importance or favoring one side over the other in a complex political controversy. Frequency measures are more likely to succeed as indicators of frequency-related phenomena—for example, the number of mentions of crime and the number of people believing crime to be an issue (not to be confused with actual crime statistics, which can be very detailed and may not correlate with public concerns), the number of favorable references to a political candidate and the number of votes that the candidate is likely to attract (not to be confused with how much the candidate has done for his or her constituency), or the proportion of unfavorable letters written to city hall (Krendel, 1970) and the likelihood that the mayor will not be reelected.

The use of Dollard and Mowrer's (1947) discomfort-relief quotient demonstrates some of the difficulties involved in establishing an index. Dollard and Mowrer applied learning theory in deriving this very simple quotient as an index of the anxiety of speakers. The quotient is computed as the proportion of the number of "discomfort" or "drive" words and the sum of this number and the number of "comfort" or "relief" words. Despite Dollard and Mowrer's sound theoretical arguments and careful definitions of the two kinds of words, tests of the indicative power of this quotient have led to mixed results. Significant correlations with palm sweating have been reported, but correlations with other measures of anxiety seem to be demonstrable only in very restricted circumstances. Murray, Auld, and White (1954) compared the discomfort-relief quotient with several other motivational and conflict measures applied during therapy and found that the quotient was not sensitive to changes in therapeutic progress. What the quotient indicates is therefore far from clear and simple.

The empirical evidence for indices of readability is more convincing. Clearly, sentences that include foreign expressions, long and compound words, complex grammatical constructions, and many punctuation marks are more difficult to read than simpler sentences. The success of Flesch's readability formula may well lie in two of its features: (a) Overall judgments concerning the readability of a piece of writing are formed cumulatively, with each encountered difficulty reducing the readability score, and (b) the indices are validated by the judgments of a population of readers. Both of these features are frequency related (Krippendorff & Bock, 2009, Chapter 3.9). Many word processing programs now are capable of providing not only counts of the numbers of characters, words, paragraphs, and pages in a document but also a readability score. Such scores might lend themselves to interesting correlational studies.

Researchers have also used indices successfully to settle disputes about authorship. In the 1940s, Yule (1944), an insurance statistician, reconsidered whether Thomas à Kempis, Jean Gerson, or one of several others wrote *The Imitation of Christ.* He correlated frequencies of nouns in works known to have been written by each prospective author and thereby developed discriminating indices to their identities, which he then applied to the disputed work (the inference was in favor of à Kempis). Mosteller and Wallace (1964), arguing that the choices of nouns are more specific to content than to author identity, found function words to be far more distinctive in their effort to settle the disputed authorship of 12 of the *Federalist Papers.* Evidence from their analysis favored Madison as the author, a finding that historians increasingly believe to be correct.

Again, declarative definitions are not sufficient. Declaring frequencies to be a measure of attention does not make them a statistical index that correlates with attention as measured by any other means. Even where correlations are found between an index and what it is said to indicate, there remains the problem of generalizability. For example, Morton and Levinson (1966) analyzed Greek texts by known authors and extracted seven discriminators of style that, according to the researchers, tap the unique elements of any person's writing: sentence length, frequency of the definite article, frequency of third-person pronouns, the aggregate of all forms of the verb *to be,* and the frequencies of the words *and, but,* and *in.* Morton's (1963) analysis of the 14 Epistles attributed to Paul in the Bible led him to conclude that six different authors wrote these works and that Paul himself wrote only four of them. Ellison (1965) then applied the constructs that Morton used to texts by known authors, which led to the inference that James Joyce's novel *Ulysses* was written by five different authors, none of whom wrote *A Portrait of the Artist as a Young Man.* Ellison found in addition that Morton's own article was written in several distinct styles. This research casts serious doubt on the generalizability of Morton's stylistic indices of an author's identity.

The inability to demonstrate high correlations should not prevent analysts from using quantitative measures, however. Researchers may be able to strengthen the indicative capabilities of such measures by adding independent variables, or they may observe these measures for long periods of time and then construct regularities that can be extrapolated into yet-unobserved domains. In addition, researchers may vindicate their construction of such measures by successfully correlating them with other phenomena not initially anticipated (correlative validity). In any case, it is always advisable to use indices cautiously.

In a self-reflective moment, Berelson (1952) wondered what Martians might infer from the high frequencies of love and sex found in modern Earth's mass-media

recordings: Would they infer a promiscuous society or a repressive one? As noted above, Pool (1952b) has observed that symbols of democracy occur less frequently where democratic processes govern than where they are in question; thus they represent something other than the degree to which democracy is accepted. Although most learning theories suggest that repetition strengthens beliefs, repetition is also known to lead to semantic satiation—not only a loss of interest but also a loss of meaning. Thus it is not a simple matter to determine what it is that frequency measures indicate, and it is certainly not an issue that can be settled by proclamation.

3.5 LINGUISTIC RE-PRESENTATIONS

In language, the analogue of indicating is *naming*. Both establish one-to-one relationships—in the case of indices, relationships between two kinds of variables, and in the case of naming, relationships between words and particular persons, things, concepts, or experiences. A name recalls the named. Although narratives use names, naming is not sufficient to allow us to understand what narratives do. Narratives conjure, bring forth, and make present (re-present as they are reread, hence *re-presentation,* with a hyphen) rich worlds consisting of people in relationships with each other: objects that do things: and ideas, morals, and perspectives that guide observations. Narratives are imaginable and, under favorable circumstances, realizable through actions. Thus, texts do not merely map, speak about, or indicate features of an existing world, they can *construct worlds* for competent speakers of a language to see, enact, and live within. To analyse *texts as re-presentations*—not to be confused with picture-like representations— *is to analyse the conceptual structure that a text invokes in particular readers,* the worlds they can imagine, make into their own, and consider real.

Written text is not just a collection of words; rather, it is sequenced discourse, a network of narratives that can be read variously. Hays (1969) provides the following examples of some typical streams of text that social or political scientists may be interested in understanding:

- *A sequence of editorials*: The staff of a newspaper, experiencing an epoch, produces a series of essays that recapitulate some of the day's events, placing them in context with respect to historical trends, theory, and dogma. The essays express opinions about the true nature of situations that are necessarily not fully comprehended and about the responses called for.

- *International exchanges of an official character:* This kind of correspondence is comparable to a sequence of newspaper editorials as described above, except that there are two or more parties involved, each pursuing its own policy.

- *Personal documents:* These may be letters, diaries, or written materials of other kinds. Such materials differ from newspaper editorials or official governmental exchanges in the particularity of their content.

- *Interview transcripts:* Usually in an interview situation there are two parties, one naive and the other sophisticated. The purpose of the interview may be, for example, therapeutic or diagnostic.

- *Social interaction:* Two or more persons participate, discussing a fixed task or whatever other topic they deem suitable.

Such streams of texts, which could be extended to include types of literature, folktales, reports of scientific findings, and corporate reports, have several characteristics in common. For instance, they are all sequential in nature. Narratives respond to each other and are no longer individual accomplishments. The structures of interest are not manifest in vocabularies of words or in sentential constructions but in larger textual units, in intertextualities. An analysis of texts as re-presentations has to acknowledge the connectedness of these larger textual units. The container metaphor that informed early conceptions of content analysis continues to influence many content analysts, making them most comfortable with classifications of content and indices that tend to ignore linguistic or narrative structures. Because such textual data tend to stem from several narrators, not one, analysts cannot presume consistency from narrator to narrator. Nevertheless, inconsistencies make sense as motivators of interactions and as causes of evolution. Re-presentations essentially provide *conceivable worlds,* spaces in which people can conceptualize reality, themselves, and others. An analysis of these re-presentations proceeds with reference to designated readers, the imaginability of actors and actions, and how each datum contributes to the unfolding of the data stream.

A simple yet generic example of such content analysis is the development of maps. Maps are not just descriptive. The user of a map needs to understand that map in order to participate in the alleged reality that the map depicts. A road map aids a driver in seeing the possibilities for realizing self-chosen goals. Without a map, the probability of the driver's reaching his or her destination would be no better than chance. But maps not only enable, they also constrain thought and enforce coordination of their users relative to each other. Inferences drawn from maps should concern what their users do or could do with them. Lynch (1965), an architect, placed verbal statements of what informants recalled seeing when moving within a city onto a composite map of that city as seen by its residents. He wanted to infer what city planners should do to provide citizens with needed orientations but found also how and where people would go when they had particular goals in mind. In his book, *Letters From Jenny,* Allport (1965) reported the results of an analysis of personal correspondence, showing what the world of its author looked like and what kind of psychological insights one could derive from her reality constructions. Gerbner and Marvanyi (1977) developed maps of the world based on their analysis of news coverage in U.S., East European, West European, Soviet, and some Third World newspapers; they distorted the sizes of the regions in the maps to correlate with the volume of news devoted to the regions. So (1995) developed maps of the field of communication research based on the titles of papers presented at several of the International Communication Association's annual conferences and on the sources cited in the papers in order to infer the "health" of the discipline. Although all the studies I have mentioned here as examples lacked good ways of tapping into complex linguistic structures, the researchers who conduct such studies tend to compensate for this shortcoming by providing rich interpretations of their findings.

Qualitative content analysts clearly recognize the need to respond to texts as connected discourse. Such researchers have examined the social construction of emotions in everyday speech (Averill, 1985), the metaphorical notion of facts in scientific discourse (Salmond, 1982), the prejudicial path toward an institutionally acceptable understanding of the causes of AIDS in medical writing (Treichler, 1988), the role of psychotherapists as depicted in fictional literature featuring psychotherapists

(Szykiersky & Raviv, 1995), the portrayal of African Americans in children's picture books in the United States (Pescosolido, Grauerholz, & Milkie, 1996), the construction of natural disasters in U.S. print media (Ploughman, 1995), and the depiction of women in the media, to name a few topics. To be clear, many of these qualitative studies have lacked formalization, and so the findings are difficult to replicate or validate. Many of these studies have also had avowedly descriptive aims; in some cases, the researchers have stated their intent to reveal biases in representations. For example, Gerbner and Marvanyi (1977) created the maps mentioned above with the intention of appealing to a fairness standard of equal attention. The use of content analysis to describe how particular media depict members of certain professions, people from certain nations, or certain social problems or political figures usually amounts to the development of maps in which the concepts of interest occupy certain places.

Analysts of re-presentations seek to rearticulate relevant portions of texts to make the readers of their analyses aware of alternative readings or readings by particular others. For example, critical discourse analysts offer accounts of the roles of language, language use, and (in)coherences and of the communicative uses of texts in the (re)production of dominance and inequalities in society (see Van Dijk, 1993). Critical discourse analysis also includes an element of self-reflexivity in that it may be applied to its own text—asking what critical analysis is, what its practitioners do to a text, and so on. Such analyses have been characterized as explorations of social cognition and the public mind. However, in the absence of the reality that re-presentations bring forth, the only criteria applicable to the analyses of re-presentations are whether they *answer* informed readers' *questions,* whether they can *withstand critical examination* from the perspective of individuals who are familiar with the context of the data, and whether *the worlds they rearticulate resemble or add to the worlds of specified readers* of the analyzed texts or of other content analysts.

Examples of analyses of re-presentations that start from the other end of this spectrum of complexity are found in simulations of cognitive processes (Abelson, 1968) and in applications of such simulations to aid political campaigns (Pool, Abelson, & Popkin, 1964). In such research, analysts use a large number of grammatically simple propositions, goals, and scripts that people know how to follow—for example, how to order a meal from a menu, how to drive a car, or how a kinship system works (Wallace, 1961)—and compute entailments from the way they hang together semantically. Without the use of computers, but certainly with that in mind, Allen (1963) proposed a logical content analysis of legal documents that demonstrated, by means of a formal procedure, which options (loopholes) existed for the signatories of an arms limitation agreement. This led Allen to infer the directions in which the parties to this agreement could, and probably would, move, given appropriate incentives, and the conflicts that could be expected to emerge. Emphasizing constraints rather than options, Newell and Simon (1956) proposed a "logic theory machine" that shows how a sequence of logical implications (a proof) from available evidence (premises, axioms) may lead to decisions within an unknown problem area (the validity of a theorem). Danowski (1993) used the data obtained from a semantic network analysis to arrive at recommendations concerning how persuasive messages ought to be constructed. Semantic network analysis is the content analysts' version of expert systems that artificial intelligence researchers aim to build in various empirical domains.

Hays (1969) developed a vision for this kind of content analysis, calling it *conversationalist*. It would accept a stream of linguistic data—dialogue, diplomatic exchanges, treaty negotiations, and the like. It would recognize that an understanding of any linguistic form presumes a great deal of background knowledge, including knowledge about beliefs and assumptions, and it would allow for such knowledge to be added to the linguistic data. If several interlocutors populate the context of an analysis, which is typical, the analysis must acknowledge differences in their background knowledge as well. The analysis would also recognize that meanings change over time and would place every assertion in the context of previous assertions. A content analysis of re-presentations, Hays's conversationalist, would *answer questions of interest to the analyst that are not literally found in the text.* The conversationalist is an engine that computes a text's implications that answer the questions given to it.

In the terms employed in our framework, as described in Chapter 2, the context of such content analyses is the reality that available texts make present to a specified community of readers. The stable relations are manifest in the reasons that the community of readers would accept for answering specific questions from specific texts, for pursuing the logical implications of these data to a chosen target. Although many content analyses of re-presentations are not so clear about their aims and rarely care to go as far, this idea is being realized, at least in part, in fifth-generation computers, so-called expert systems. The discussion of expert systems has been overshadowed by interest in search engines for the Internet, computer networking, and collaborative systems, to name just a few, but the fact that we now have large volumes of textual data available in computer-readable form makes the content analysis of re-presentations increasingly possible and a challenge.

Sherlock Holmes's detective work, already mentioned in section 2.4.5, provides a literary example of the use of linguistic re-presentations. Content analysts merely deal with larger volumes of textual matter, pursue different kinds of questions, and, for the process to be examinable or replicable, need to proceed transparently.

In this context, I might mention various visualizations of less directly language-related textual relationships such as webgraphs, exemplified by Figures 11.5 and 11.6. Webgraphs depict the links between one website and another, between a source URL and a destination URL. They exemplify a hypertextual conception of text that content analysts have to cope with. Ordinary text is expected to be read as written. Hypertext gives the reader far more freedom to navigate a textual multiverse, and gives content analysts considerable headaches for how to account for the many ways a hypertextual multiverse can be and is, in fact, visited. While the use of URLs in websites is a technological modification of the traditionally linear conception of text, the networking of what is said is ancient, of course. It may not always be recognized. However, I would say that everything said is addressed to someone; and when an addressee listens and responds to what is heard, what is said, the text, has established a relationship. This is what the literary theorist Mikhail Bakhtin (Todorov, 1988) taught in opposition to the linguistic conception of language—which blatantly omits relational, social, and dynamic conceptions of texts, moving from reader to reader, being transformed in that process, but establishing, maintaining, and replacing networks of text-based relationships. This is a good introduction to the next section.

3.6 CONVERSATIONS

When children in well-to-do families say they are hungry, they may well want to have something to eat, but they may just as well be trying to avoid going to bed, gain attention, prevent their parents from doing something, or so on. In the context of a lived history of interacting with their children (knowing when they last ate, for example), parents tend to know how to respond when their children claim to be hungry. In such a situation, the propositional content of an utterance is secondary to the role that utterance plays in an ongoing interaction. In an attempt to infer anxiety from speech, Mahl (1959) addressed the difficulties of analyzing this kind of instrumental use of language, but he ended up bypassing the problem in favor of developing nonverbal indicators of anxiety instead. The path he took demonstrates the limitation of content analyses that are guided by a representational concept of content. Already in the 1950s, Bateson (2000; Ruesch & Bateson, 1951) had suggested that all messages convey content *and* relational information (a concept addressed by many researchers since, from Watzlawick, Beavin, & Jackson, 1967, to Baxter & Montgomery, 1996). When we view utterances as only representations, we ignore their relational or conversational functions. The essential feature of *conversational interactions* is that they *take place in and create interpersonal relations* and *define their own conditions for continuing the process.* When we blame someone for lying, we invoke the standard of representational truths, which is only one of many possible conversational frames interlocutors can adopt and one that makes continuing a conversation less important than being right. In content analyses of conversations, *inferences concern the continuation of the process.* Indexical and re-presentational aspects (content in the sense of what is conveyed in processes of communication) are at best a means to that end.

Conversation analysis has emerged as one approach to the study of talk in natural settings (Atkinson & Heritage, 1984; Goodwin, 1981; Hopper, Koch, & Mandelbaum, 1986; Jefferson, 1978; Sacks, 1974; ten Have, 1999). Unlike discourse analysts, who start with written texts (Van Dijk, 1977, 1993), regard a discourse as a string of sentences, and aim to account for what the discourse (re)presents, as well as how and why it (re)presents what it does, conversation analysts tend to start with voice or video recordings of naturally occurring speech. They then proceed by transcribing conversational interactions, using highly specialized transcription conventions that enable them to capture not only words and who uttered them but also intonations, overlaps, and incompletions, as well as nonverbal behaviors such as gaze and especially silences and turns at talk. Broadly speaking, conversation analysts aim to understand the structure of naturally occurring speech, which necessarily includes two or more of its participants. Their methods of study are intended to preserve as much of the richness of human communication as possible. One typical analytical strategy is to differentiate among speech acts, or utterances that do something, such as questions, requests, promises, declarations, and expressions of feelings that are constitutive of relationships between the conversants.

Although conversation analysts are beginning to address reliability issues in their studies (e.g., Carletta et al., 1997; Patterson, Neupauer, Burant, Koehn, & Reed, 1996), efforts to establish the validity of conversation analyses have been marred by a lack of consensus concerning what constitutes supporting evidence. Most published

reports of conversation analysis research can be characterized as "show and tell." In these reports, researchers reproduce exemplary fractions of transcribed dialogue to demonstrate their explanations of "what is 'really' going on." It is generally futile to ask conversants to confirm conversation analysts' claims, as ordinary speakers engage each other "on the fly" and without access to or understanding of the analytical tools that conversation analysts have developed to transcribe and examine verbal interactions in great detail. However, inasmuch as conversations involve several participants whose utterances are made in response to previous utterances and in anticipation of future responses (thus the process is directed from within a conversation), researchers have the opportunity to understand conversations as cooperatively emerging structures that are, at each point in the process, responsive to past interactions and anticipatory of moves to come. A content analysis of data as conversation could involve (a) *inferring from any one moment of a recorded history of interactions the range of moves that could follow,* (b) *reinterpreting that history from the moves that actually did follow,* and (c) *systematically applying this explanatory strategy to all moments of naturally occurring conversations.*

This form of analysis is applicable not just to everyday conversations but also to exchanges between actors in organizational roles or as representatives of national governments. In exchanges between managers and employees, just as between therapists and their clients or between professors and their students, power issues enter through the speech acts the interlocutors choose, accept, or deny each other. Power relationships have become a favorite topic of critical scholarship among conversation analysts. Social organizations can be seen as reproducing their members' commitment to the preservation of the organizations' form. Commitments need to be asserted, heard, believed, and enforced. Thus, organizations reside in certain speech acts, in how members respond to each other's talk. This makes organizations analyzable as networks of conversations of a certain kind. Analyses of exchanges between representatives of nations are not new, but conversation analyses of the unfolding dynamics in such exchanges offer a new approach to international relations. Content analyses of negotiations have advanced an understanding of the process (Harris, 1996). Pathologies of communication gain new currency when analyses reveal restrictions or constraints on conversation. Some scholars have called for the quantification of interactions (e.g., Hopper et al., 1986). This has been demonstrated, for example, regarding doctor-patient interactions (Ford, Fallowfield, & Lewis, 1996; Katz, Gurevitch, Peled, & Danet, 1969).

3.7 INSTITUTIONAL PROCESSES

The foregoing discussion has moved beyond the traditional notion of content in three senses: in the assertion that content analysis may be applied to any kind of data—texts, images, sounds, music, artifacts, anything that humans are able to vary for effect or unconsciously reproduce, in the assertion that analysts may draw inferences from textual data to features of any specifiable context, and in the assertion that texts are part of relational dynamics, of networks of relationships. In this section, I discuss a further expansion of the scope of content analysis to include inferences about institutional phenomena of which the institutions' constituents may be only dimly aware. Much communication that takes place within institutions is routine,

relational, and coordinative, and valued as such, even enforced, without apparent reason. Moreover, institutions reside in particular qualities of communication. Because communication in institutions tends to go beyond unaided readers' scope of comprehension, content analyses that probe into institutional properties call for *analytical instruments and theories* that, like microscopes, telescopes, and computer intelligence, *provide inferential access to social realities that are too complex to be accessible otherwise.*

Berger and Luckmann (1966) outline the context of this kind of content analysis. To start, they suggest that *habitualization* is an important prerequisite of institutionalization:

> Any action that is repeated frequently becomes cast into a pattern, which can then be reproduced with an economy of effort and which, *ipso facto,* is apprehended by its performer *as* that pattern. Habitualization further implies that the action in question may be performed again . . . with the same (or reduced) economical effort. . . . Habitualization carries with it the important psychological gain that choices are narrowed. (p. 53)

An example is the pattern of grammar, which directs our thoughts and actions in ways we rarely notice. For instance, the English language recognizes just two genders. Consequently, native English speakers tend to find the distinction between males and females natural and obvious. In turn, many institutions in English-speaking societies are built on this distinction. Gender is not a given phenomenon. It is an artifact of enacting grammar and vocabulary, which change only slowly. Because of the linguistic distinctions we confuse with facts, numerous problems arise, for example, by providing little if any space between the two gender identities. Such problems manifest themselves in gay bashing, the shunning of transvestites, and the difficulties that androgynous people face in their lives. That members of other cultures draw different distinctions demonstrates the institutional nature of such grammatical categories.

The ways in which we greet each other every day, the repetitive and utterly predictable categories of television programming and news coverage that we have come to take for granted, the ceremonial nature of the political process, the pervasiveness of climates of power in everyday life—all such patterns, weathered in the process of apparent successes, are the backbone of the institutionalization of human behavior. However, the comfortable certainties that this kind of habitualization offers also suppress our ability to see untested alternatives. Content analysts who study institutions can infer habitualization from repetition and the narrowing of choices from the absence of mention of alternative ways of being or doing things. Surprisingly, Shannon's information theoretical notions of redundancy (a quantification of the nonuse of otherwise available alternatives) and information (a measure of surprise in the context of available messages) can be seen to have institutional interpretations (see Shannon & Weaver, 1949).

Regarding habitual *patterns,* to the extent that people are concerned, they talk of them in a language that *categorizes* (typifies) not only the *actions* that constitute these patterns but also the *actors/participants* involved. Teaching is what teachers do in front of students. Entertaining is what entertainers do for their audiences. Such obvious and semantically tautological propositions involve categories that furnish people with spaces they may occupy or grow into, not only to fit the categories but also *to see*

each other in terms of those categories. Berger and Luckmann (1966, p. 54) identify institutions in terms of such *reciprocal categorizations.* For instance, knowing what a bank is enables the individuals in the context of a bank to interact with each other as certain categories of people—customers, tellers, guards, bank presidents, even bank robbers—regardless of who they are as individuals and regardless of whether the bank is in Philadelphia or Hong Kong. The same applies to understanding what is being said as proper. Reciprocal categorizations provide a key to how institutions are constituted, and such categorizations are easily identified in the texts that institutions generate—employee handbooks, memos on hiring practices, minutes of meetings, reports to shareholders—especially those generated by mass-media entertainment, which is highly institutionalized.

Berger and Luckmann (1966) note that *we grow into a world already constructed by others,* without knowing why things have become the way they now are. This lack of knowledge of the histories of institutions leads to the belief that "things are good because this is how they have 'always' been." Such a belief discourages, if not punishes, deviations from established patterns. Thus, our lack of knowledge of history fuels institutional controls. Moreover, playing down human participation, without which institutions cannot exist, we tend to transfer agency to institutions, as when we say, "Science says . . . ," "The media show . . . ," or "The military discriminates against . . ." We consider institutions to be capable of preserving themselves, as when we speak of the "interests of government." Institutions do not really control deviance from institutional patterns, nor do they assign powers to the roles people play in them. Individual participants police each other. Talk and texts that prevents people from realizing the roles they play in maintaining institutional practices is one target of content analyses of institutions.

Institutions tend to remain hidden behind habitual practices until flaws emerge and certainties break down. Families, bureaucracies, and nations are unthinkable without routine forms of communication. Family therapy is justified by the experience of routines whose mindless enactment within a family gets their members into trouble. Most institutions are taken for granted. Whereas the culturally dominant individualist approach to therapy encourages families to identify one family member as the one who needs help, working with that person as the identified patient rarely changes the dynamics of interactions within the family, perpetuating its dysfunctionality.

The emergence of international conflicts, which are rarely ever intended and in which nobody really likes to participate, especially when they may die as a result, is evidence of the fact that people can get involved in such events without knowing how. It thus appears that the way individuals create and preserve the institutional structures they live with are barely recognized and are given free range to play out. Institutions are not transparent unless we witness moments of breakdown, when something gets its members into unknown trouble in need of explanation, fixing, or trying to escape from its hold on them and are prevented from doing so. Content analyses of institutions often focus on communications at moments of such breakdowns. For instance, Berelson's (1949) study of "what 'missing the newspaper' means," conducted during a newspaper publishing strike in New York, revealed numerous previously unrecognized uses of newspapers and how their absence made people aware of these institutionalizations.

It is at moments of emerging difficulty or fear of the consequences of deviation from normalcy that the human constituents of institutions invent, appeal to, and apply institutional rules of conduct. Often such rules grow into systems of institutionalized explanations that become accessible through the very communications that invoke them. Mass communications researchers who have examined institutional processes have focused on legal, economic, political, technological, that is, structural explanations for those processes.

Legal explanations emphasize that communicators operate under certain legal conditions; for example, they may have to be licensed or must comply with or avoid violating contractual arrangements. Communicators may speak in an official capacity for particular social organizations, or they may question the legitimacy of certain practices. Texts obtained in legally regulated contexts reflect the legal constraints under which the institutional communicators who are being studied operate.

Economic explanations emphasize that when communication (production, transmission, and consumption) incurs costs, someone must pay them in some form, which creates networks of stakeholders with economic interests in what is being communicated. For example, in the United States, the traditional mass media are paid for largely by advertising; thus what is aired must be profitable in some way and cannot offend its sponsor. The effects of media ownership—in particular, the effects of monopolies and certain commercial interests—on communications have been a frequent target of content analyses.

Political explanations emerge when communications are disseminated widely and become of concern to competing public interests. The institution of ethical standards may result from debates about apparent problems, such as the violation of individuals' privacy by photographers (paparazzi) or reporters in the competitive pursuit of sensationalist material. Intentional misrepresentations by journalists and campaign advertisements perceived as "dirty," "slanderous," or "unfair" may lead to legal actions. Newspaper publishers, television stations, and other kinds of organizations cannot afford to displease vocal publics if they expect to have some longevity, whether the publics that concern them are made up of governing elites or masses of audience members. Thus, communications reflect as well as enact the prevailing configurations of rhetorical power. In conducting content analyses aimed at examining power relationships, researchers have to be careful not to fall into the trap of believing that everyone perceives power as they do. Instead, they may want to look at how power is enacted, received, or undone (Krippendorff, 1995b).

Technological explanations of institutional processes emphasize that all communications must be producible, recordable, distributable, and accessible through various technological means, and that actual communications are not only shaped in that process but also shape the institutions in whose terms they are processed. The film and television industries employ techniques of mass production that are vastly different from those employed by newspapers. This is a matter not of intention but of the nexus between technologies and the institutions that thrive on them. An even greater difference exists between newspapers and what computer-mediated communication—the Internet, for example—can make available. Content analyses have shed light on the systematic changes in content that take place when a book is made into a film (Asheim, 1950), on the role of gatekeepers in news flow (e.g., what happens to controversial content; White, 1964), on how news is made

as opposed to reported (Gieber, 1964), on the social role of the magazine cover girl as a function of channels of distribution (Gerbner, 1958), and on how expectations about institutions shape the forms of petitions directed to those institutions (Katz, Gurevitch, Danet, & Peled, 1969). In addition, based on an examination of mass communication from a technological perspective, Adorno (1960) has contended that the institutionalized repetitiveness of the mass production of messages preserves and strengthens social stereotypes, prejudices, and ideologies rather than corrects them (see Section 3.2.3).

There are a few fundamental generalizations from which content analyses of institutionalized texts may start. One is that everything said, written, listened to, or read—every communication—not only says something to its receiver but also institutes the very pattern of which it is a part. For example, a person who cashes a check at a bank is not merely taking part in a mutually beneficial transaction; his or her action also manifests trust in money and supports banking as an institution. If people did not bank regularly, the banking industry could not exist. Banks are instituted in our trust in money, in our belief in the safety of banks, in the belief that one should earn interest on savings, and so on. When people turn on their television sets to see certain shows, they are not only gaining entertainment but also supporting the shows they watch by increasing the programs' ratings. Their actions also legitimate the mass media as an institution to provide such entertainment. If nobody were to watch television for a while, the mass media could not exist as usual. The use of racial categories—whether on television, in everyday talk, or in survey questions—demonstrates that they are important, of public or interpersonal concern, and the very use of these categories invariably strengthens ethnic prejudices and makes them real. People's participation in a pattern of reciprocal categorization is an essential requirement for institutions to persist, and this applies also to issues of race. Therefore, analyses of such communication phenomena cannot stop at what is being said or heard. What matters is that the very act of communication strengthens that act, allows for repetition, and keeps people in attendance. Researchers conducting content analyses of institutionalized texts—which most mass communications are—have to observe whether communications constitute new patterns, strengthen what has been said before through repetition, or weaken a pattern by omission or attention to alternatives.

The properties of the medium of communication in which an institution is constituted have profound effects on the development of that institution. Innis (1951) compared oral communications with written communications and concluded that writing has the effect of freezing traditions, rendering institutions more permanent and reliable; thus written communications can support empires that extend control over larger geographic areas. Radio and television, with their virtually instantaneous transmission over vast distances, tend to support the development of geographically widely dispersed organizational forms, but because such media do not leave many records behind, these forms are far less stable than those supported by written media. Oral and audiovisual media are also less controllable than written ones. The revolution against the shah of Iran succeeded largely because of the support generated among the people of Iran through the distribution of audiotapes that escaped the control of the state, unlike all other media in that country. The short-lived pro-democracy movement in China was organized largely through fax communication, which recognized no national boundaries. The Internet now provides vast numbers

of geographically dispersed users with almost instantaneous access to computer-mediated communications; moreover, it is capable not just of disseminating data but of facilitating discussion among parties to negotiations and commitments as well as the rise of commercial enterprises and virtual communities that can undermine more traditional institutions. These diverse spatial, memory-related, and coordinative properties of communications media have profound effects on institutional dynamics, and research that focuses on single messages or on readings of individual users cannot possibly reveal those effects. Content analyses in institutional contexts can lead to inferences regarding the weakening or strengthening of certain institutions, and frequencies play important roles in such inferences.

Finally, *communications tend to reinforce the very institutional explanations and rules by which they are created and disseminated.* For one example, the traditional mass media, which operate in conformity with the one-way theory of communication, produce communications that demonstrate the workings of this theory and are likely to encourage the widespread use of this theory at the expense of alternative ways of thinking about human communication. It is therefore not surprising that from its inception, the field of communication research has been fundamentally committed to a model of communication that consists of a sender, messages, and receivers—as if no others, bystanders and judges, were relevant and worthy of attention. One-way communication technology has given rise to totalitarian regimes but also to disciplines such as advertising and causal theories such as of persuasion, influence, and effects. The emergence of interactive media has challenged this paradigm somewhat, but even today many researchers who seek to understand computer-mediated communication, which is well instituted, start with mass communication models. To infer institutional controls, content analysts have to observe what is not said, what happens when institutions are challenged or break down, and what is done to those who dare to deviate from institutionalized practices. Content analyses of textual data in view of such phenomena can add to our understanding of the workings of institutionalized patterns of thinking and acting. All of these phenomena become evident through analyses of the use of language.

Lasswell (1960) sought to clarify the institutional roles of communication by distinguishing among three functions:

- Surveillance of the environment
- Correlation (coordination) of a society's parts in response to the environment
- Transmission of social heritage from one generation to the next (culture)

To these, Wright (1964) added one more function:

- Entertainment

Echoing Parsons's (1951) sociological systems theory, both Lasswell and Wright argue that any society has to develop institutions that specialize in performing these functions. In the United States, journalism could be seen as serving the surveillance function in that journalists report publicly on events that take place, and politics could be regarded as serving the correlation function by mobilizing individuals to behave in ways that serve society as a whole and coordinating the distribution of resources (Lasswell, 1963). Functionalist accounts of institutions, to which this classification

is indebted, are not the only accounts, however. Nor can they be accepted without question, for they preserve sociological theorists' vision of what society is. Instead of imposing theoretical categories from outside or from a position of authority, content analysts attempt to understand institutions through how the participants in them talk with each other and about their own institutional involvements, how they participate in maintaining these institutions through talk and writing, and how they judge the legitimacy and appropriateness of institutions' actions. Institutions are constituted, constructed, and reconstructed in language use and in the distribution of narratives of that use through particular media of communication; hence the need for content analyses of this kind. Lasswell's classification at least acknowledges that institutions are manifest in all communicative practices and serve as the glue that holds a society together.

3.8 AREAS OF LIKELY SUCCESS

Having completed the above survey of areas in which content analysis may be applied, I conclude this chapter by offering a brief answer to the question of where content analysis might be used most fruitfully. To this end, I draw on the conceptual issues discussed in Chapter 2, including my argument against the use of the container metaphor, which entails the conception of content as a tangible entity, contained in messages and shipped from one place to another, that researchers presume to be able to analyze through the use of objective (i.e., observer-independent) techniques. I have argued instead for the metaphor of reading, which shifts attention from what content is to what readers do with texts, how they relate texts to the contexts of their use—individually, politically, socially, and culturally—and what this means for various social phenomena. In this shift, readers and their communities become central, whether the readers are authors, users, bystanders, or content analysts. Content analysts cannot exclude themselves from the population of readers, albeit their reading is aided by systematic methods of careful inquiry.

To determine where content analysis is most fruitfully employed, we should consider texts as the by-products of ongoing conversations. We have to acknowledge that people learn to read and write a language only after they have learned to speak it, that texts build on the experience of speech. They can substitute for speech, as when people write letters. They can extend the range of speech, as in the use of communication technology—mass and computer-mediated communication— and they usually revert back to speech, by being read, interpreted, talked about, accepted, dismissed, or selectively enacted. Texts are more durable than speech. Texts may be reread in support of gaining new insights, and they may be analyzed repeatedly and replicated by several analysts. This is why conversation analysts, for example, record and transcribe speech before they analyze it or talk of it. Whether texts are written for fear of forgetting something; to assure their writers desirable certainties, as when signing a contract; or to grant certainties to others, as when writers delegate their agency to a lawyer or government to act on their behalf, most writing shifts some kind of authority. Literacy, the ability to read and act on a text like everyone else, is expected by all members of a speech community. This stability is questionable, of course, as are the institutions whose members are dedicated to preserve them.

From my perspective, texts are always rooted in the dialogical context of conversations, subject to adaptation to changing situations, open to reinterpretations in dialogue among community members, even if such changes remain unnoticed. Content analysts may be able to chart these changes, even if they are not recognizable by those who reproduce the interpretations of texts.

Scholars have studied the effects of reading and writing, and the uses of the mass media, for some time, although this research has always lagged behind technological developments (information technology, for example). As noted above, typical topics of research have included deception, attitude change, message effects, uses and gratifications, technological biases, rational decision making, institutionalization, and causal connections between textual and nontextual phenomena.

Regarding assessments of crime, unemployment, and the economy, for example, studies have repeatedly demonstrated that correlations between what the mass media present—text in our generalized sense—and what public opinion polls find or individuals express as concerns are higher than the correlations between either of these and actual statistics on crime, unemployment, and the economy. This suggests that content analyses are generally more successful when they stay close to the uses of language—after all, public and individual opinions involve talk, not physical measurement. Another example is the well-studied phenomenon of agenda setting by the mass media (McCombs & Shaw, 1972; McCombs, Shaw, & Weaver, 1997)—that is, the fact that themes and issues distributed by the mass media in the form of print, speech, and images have a good chance of becoming topics of public conversations and, in turn, affect civic actions, inform political decisions, and stimulate artistic rearticulations. The simple reason for this phenomenon is that widely distributed texts enter, are adopted into, and come alive in conversations, not only in conversations between interviewers and interviewees but also conversations in public places ranging from side-street cafés to political demonstrations. If new words and expressions resonate with readers' or listeners' previously acquired language habits, they may take hold in the public imagination and become part of many people's vocabularies.

Content analyses are most successful when they focus on facts that are constituted in language, in the uses of the very texts that the content analysts are analyzing. Such linguistically constituted facts can be broken down into four classes:

- *Attributions*: Concepts, attitudes, beliefs, intentions, emotions, mental states, and cognitive processes ultimately manifest themselves in the verbal attributes of behavior. They are not observable as such. The words that make them real are acquired, largely in conversations but also through reading and attending to various media of communication. The attribution of competence, character, morality, success, and belongingness to particular categories of people enables or discourages actions, makes or breaks politicians, creates heroes and demonizes villains, identifies leaders and marginalizes minorities. These facts cannot exist without language, and to the extent that texts are instrumental in disseminating and creating such attributions, they are natural targets of successful content analyses.

- *Social relationships:* Noting that statements or questions can be uttered either subserviently or authoritatively, Bateson introduced the distinction between the content and the relationship aspect of all human communication

(Ruesch & Bateson, 1951, pp. 179–181). Relationships may be established implicitly or taken for granted in how communication takes place. For example, by offering third-person plural accounts of observed "Others," scientific observers set themselves apart from their subjects and assume a position of superiority. This is manifest in the grammar of talk or writing. But relationships may also be negotiated, unilaterally imposed, and explicitly accepted or rejected. Authority, power (Hillman, 1995; Krippendorff, 1995b), contractual agreements, and inequalities are all constituted primarily in *how* language is used and only secondarily in *what* is said. Content analyses tend to be more successful when they focus on how language is used, relying on social grammars of recorded speech or written communication of which speakers or writers may not be fully aware.

- *Public behaviors:* Individuals' values, dispositions, conceptions of the world, and commitments to their way of being surface in conversations that involve repeated confirmations. Without such repetition, individuals drift apart, their behaviors become no longer coordinated, and they experience difficulties in understanding each other. To the extent behavior is public, and hence observed and judged by others, it is brought into the domain of language. Narratives too, are essentially public. They may inspire individuals to act, but they are always told by someone and listened to by others, rendering inspiration a social experience. Reading a newspaper may be an individual act, but not only do newspapers print what editors consider to be of public interest, newspaper readers also talk to others about what they read, and so make newspaper reading a public activity. The vocabularies we use are all acquired from others who have used the words before us. Inasmuch as a vocabulary suggests the range of what a person can talk about and conceive, the conceivable is transmitted from parents to children, from speakers to listeners, and from writers to readers. All uses of language ultimately are public—not shared but in the open. Content analyses are more likely to succeed when they address phenomena that are of a public, social, or political nature or concern phenomena of individuals' participation in public, social, or political affairs. Cognition, for example, the supposed crown of individualism, is never an exclusively individual phenomenon. It always reflects the contexts of others, much as texts do.

- *Institutional realities:* We often overlook the institutional nature of social realities—of marriage, money, government, history, illness, and even scientific pursuits. Public opinion, for example, is a construction that relies heavily on the language of social science, on statistics in particular, but it also depends crucially on being taken as a political reality and acted upon. Without the institution of free speech, the authority of journalism, and constitutional democracy, public opinion research would not make much sense. Mental illness has an institutional reality as well. It is projected onto identified patients in terms of categories that mental health professionals and insurance companies have developed for their convenience. The factuality of these phenomena derives from certain institutionalized texts, such as the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV-R*),

which is published by the professional authority on mental illness, the American Psychiatric Association (2013). These texts legitimate numerous therapeutic interventions. For still another example, consider how a social organization such as a family or a corporation constitutes itself. Members of an organization coordinate their activities through communication and continually affirm their membership, often shielding vital stories about inside practices from outsiders. When exchanges within organizations take place in written form, they stabilize organizational memories, identities, and practices. Disrupting an organization's network of communication can cause the organization to collapse. Organizational communication research has successfully inquired into how organizations arise in the communications among members and develop nourishing organizational cultures. Content analysis of what is said and written within an organization provides the key to understanding that organization's reality, but it is most likely to succeed if it considers the more stable categories in which the organization constitutes itself.

In sum, content analyses are most likely to succeed when analysts relate textual matter—understood quite broadly to include images and traffic on the Internet—to social realities that enact them in their daily life, in conversations that yield consensus on how they are to be read or should no longer be read, that is, in conversations that keep texts alive. Repetitive, routine, public, and institutionalized phenomena are easier to infer than are rare and unconventional ones. Moreover, because content analysis presupposes familiarity with, if not literacy in, the language of the analyzed texts, the more cognizant content analysts are of vocabulary and subtle discursive conventions, including their own, the more they are able to reflect on what they do with texts that had or still have a life for others and the more likely they are to succeed and leave their contributions for everyone to read and interpret.