(Bulmer, 1979). Grounded theory, schema analysis, ethnographic decision modeling, and analytic induction all include model-building phases.

Once a model starts to take shape, the researcher looks for negative cases—cases that don't fit the model. Negative cases either disconfirm parts of a model or suggest new connections that need to be made. In either instance, negative cases need to be accommodated. Negative case analysis is discussed in detail by Becker, Geer, Hughes, and Strauss (1961, pp. 37-45), Strauss and Corbin (1990, pp. 108-109), Lincoln and Guba (1985, pp. 309-313), Dey (1993, pp. 226-233), Miles and Huberman (1994, p. 271), and Becker (1998), and is used by schema analysts (Quinn, 1997), ethnographic decision modelers (Gladwin, 1989), and scholars who use analytic induction (Bloor, 1976; Cressey, 1953/1971; Lindesmith, 1947/1968).

In ethnographic decision modeling and in classical content analysis, models are built on one set of data and tested on another. In their original formulation, Glaser and Strauss (1967) emphasized that building grounded theory models is a step in the research process and that models need to be validated. Grounded theorists and schema analysts today are more likely to validate their models by seeking confirmation from expert informants than by analyzing a second set of data. For example, Kearney, Murphy, and Rosenbaum (1994) checked the validity of their model of crack mothers' experiences by presenting it to knowledgeable respondents who were familiar with the research.

Regardless of the kind of reliability and validity checks, models are simplifications of reality. They can be made more or less complicated and may capture all or only a portion of the variance in a given set of data. It is up to the investigator and his or her peers to decide how much a particular model is supposed to describe.

Below we review some of the most common methods researchers use to analyze blocks of texts. These include grounded theory, schema analysis, classical content analysis, content dictionaries, analytic induction, and ethnographic decision tree analysis.

Grounded Theory

Grounded theorists want to understand people's experiences in as rigorous and detailed a manner as possible. They want to identify categories and concepts that emerge from text and link these concepts into

substantive and formal theories. The original formulation of the method (Glaser & Strauss, 1967) is still useful, but later works are easier to read and more practical (Charmaz, 1990; Lincoln & Guba, 1985; Lonkila, 1995; Strauss, 1987). Strauss and Corbin (1990), Dey (1993), and Becker (1998) provide especially useful guidance. (For some recent examples of grounded theory research, see Hunt & Ropo, 1995; Irurita, 1996; Kearney et al., 1994; Kearney, Murphy, Irwin, & Rosenbaum, 1995; Sohier, 1993; Strauss & Corbin, 1997; Wilson & Hutchinson, 1996; Wright, 1997.)

Grounded theory is an iterative process by which the analyst becomes more and more "grounded" in the data and develops increasingly richer concepts and models of how the phenomenon being studied really works. To do this, the grounded theorist collects verbatim transcripts of interviews and reads through a small sample of text (usually line by line). Sandelowski (1995a) observes that analysis of texts begins with proofreading the material and simply underlining key phrases "because they make some as yet inchoate sense" (p. 373). In a process called "open coding," the investigator identifies potential themes by pulling together real examples from the text (Agar, 1996; Bernard, 1994; Bogdan & Biklen, 1992; Lincoln & Guba, 1985; Lofland & Lofland, 1995; Strauss & Corbin, 1990; Taylor & Bogdan, 1984). Identifying the categories and terms used by informants themselves is called "*in vivo* coding" (Strauss & Corbin, 1990). As grounded theorists develop their concepts and categories, they often decide they need to gather more data from informants.

As coding categories emerge, the investigator links them together in theoretical models. One technique is to compare and contrast themes and concepts. When, why, and under what conditions do these themes occur in the text? Glazer and Strauss (1967, pp. 101-116) refer to this as the "constant comparison method," and it is similar to the contrast questions Spradley (1979, pp. 160-172) suggests researchers ask informants. (For other good descriptions of the comparison method, see Glaser, 1978, pp. 56-72; Strauss & Corbin, 1990, pp. 84-95.)

Another useful tool for building theoretical models is the conditional matrix described by Strauss and Corbin (1990, pp. 158-175). The conditional matrix is a set of concentric circles, each level corresponding to a different unit of influence. At the center are actions and interactions; the outer rings represent international and national concerns, and the inner rings represent individual and small group influences on action. The

matrix is designed to help investigators to be more sensitive to conditions, actions/interactions, and consequences of a phenomenon and to order these conditions and consequences into theories.

Memoing is one of the principal techniques for recording relationships among themes. Strauss and Corbin (1990, pp. 18, 73-74, 109-129, 197-219) discuss three kinds of memos: code notes, theory notes, and operational notes. Code notes describe the concepts that are being discovered in "the discovery of grounded theory." In theory notes, the researcher tries to summarize his or her ideas about what is going on in the text. Operational notes are about practical matters.

Once a model starts to take shape, the researcher uses negative case analysis to identify problems and make appropriate revisions. The end results of grounded theory are often displayed through the presentation of segments of text—verbatim quotes from informants—as exemplars of concepts and theories. These illustrations may be prototypical examples of central tendencies or they may represent exceptions to the norm. Grounded theory researchers also display their theoretical results in maps of the major categories and the relationships among them (Kearney et al., 1995; Miles & Huberman, 1994, pp. 134-137). These "concept maps" are similar to the personal semantic networks described by Leinhardt (1987, 1989), Strauss (1992), and D'Andrade (1991) (see below).

Schema Analysis

Schema analysis combines elements of the linguistic and sociological traditions. It is based on the idea that people must use cognitive simplifications to help make sense of the complex information to which they are constantly exposed (Casson, 1983, p. 430). Schank and Abelson (1977) postulate that schemata—or scripts, as they call them—enable culturally skilled people to fill in details of a story or event. It is, says Wodak (1992, p. 525), our schemata that lead us to interpret Mona Lisa's smile as evidence of her perplexity or her desperation.

From a methodological view, schema analysis is similar to grounded theory. Both begin with a careful reading of verbatim texts and seek to discover and link themes into theoretical models. In a series of articles, Quinn (1982, 1987, 1992, 1996, 1997) has analyzed hundreds of hours of interviews to discover concepts underlying American marriage and to show how these concepts are tied together. Quinn's (1997) method is to

"exploit clues in ordinary discourse for what they tell us about shared cognition—to glean what people must have in mind in order to say the things they do" (p. 140). She begins by looking at patterns of speech and the repetition of key words and phrases, paying particular attention to informants' use of metaphors and the commonalities in their reasoning about marriage. Quinn found that the hundreds of metaphors in her corpus of texts fit into just eight linked classes, which she calls lastingness, sharedness, compatibility, mutual benefit, difficulty, effort, success (or failure), and risk of failure.

Metaphors and proverbs are not the only linguistic features used to infer meaning from text. D'Andrade (1991) notes that "perhaps the simplest and most direct indication of schematic organization in naturalistic discourse is the repetition of associative linkages" (p. 294). He observes that "indeed, anyone who has listened to long stretches of talk—whether generated by a friend, spouse, workmate, informant, or patient—knows how frequently people circle through the same network of ideas" (p. 287).

In a study of blue-collar workers in Rhode Island, Claudia Strauss (1992) refers to these ideas as "personal semantic networks." She describes such a network from one of her informants. On rereading her intensive interviews with one of the workers, Strauss found that her informant repeatedly referred to ideas associated with greed, money, businessmen, siblings, and "being different." She displays the relationships among these ideas by writing the concepts on a page of paper and connecting them with lines and explanations.

Price (1987) observes that when people tell stories, they assume that their listeners share with them many assumptions about how the world works, and so they leave out information that "everyone knows." Thus she looks for what is *not* said in order to identify underlying cultural assumptions (p. 314).

For more examples of the search for cultural schemata in texts, see Holland's (1985) study of the reasoning that Americans apply to interpersonal problems, Kempton's (1987) study of ordinary Americans' theories of home heat control, Claudia Strauss's (1997) study of what chemical plant workers and their neighbors think about the free enterprise system, and Agar and Hobbs's (1985) analysis of how an informant became a burglar. We next turn to the two other methods used across the social sciences for analyzing text: classical content analysis and content dictionaries.

Displaying Concepts and Models

Visual displays are an important part of qualitative analysis. Selecting key quotes as exemplars, building matrices or forms, and laying theories out in the form of flowcharts or maps are all potent ways to communicate ideas visually to others. Models are typically displayed using boxes and arrows, with the boxes containing themes and the arrows representing the relationships among them. Lines can be unidirectional or bidirectional. For example, taxonomies are models in which the lines represent the super- and subordinate relationships among items. Relationships can include causality, association, choices, and time, to name a few.

A widely used method for describing themes is the presentation of direct quotes from respondents—quotes that lead the reader to understand quickly what it may have taken the researcher months or years to figure out. The researcher chooses segments of text—verbatim quotes from respondents—as exemplars of concepts, of theories, and of negative cases. Ryan (1999) has used multiple coders to identify typical quotes. He asks 10 coders to mark the same corpus of text for three themes. Ryan argues that the text marked by all the coders represents the central tendency or typical examples of the abstract constructs, whereas text marked by only some of the coders represents less typical examples and is more typical of the "edges" of the construct.

Tables can be used to organize and display raw text or can be used to summarize qualitative data along multiple dimensions (rows and columns). The cells can be filled with verbatim quotes (Bernard & Ashton-Voyoucalos, 1976; Leinhardt & Smith, 1985, p. 254; Miles & Huberman, 1994, p. 130), summary statements (Yoder, 1995), or symbols (Fjellman & Gladwin, 1985; Van Maanen, Miller, & Johnson, 1982). (For a range of presentation formats, see Bernard, 1994; Miles & Huberman, 1994; Werner & Schoepfle, 1987.)

Classical Content Analysis

Whereas grounded theory is concerned with the discovery of data-induced hypotheses, classical content analysis comprises techniques for reducing texts to a unit-by-variable matrix and analyzing that matrix quantitatively to test hypotheses. The researcher can produce a matrix by applying a set of codes to a set of qualitative data (including written texts as well as audio and video media). Unlike grounded theory or schema

analysis, content analysis assumes that the codes of interest have already been discovered and described.

Once the researcher has selected a sample of texts, the next step in classical content analysis is to code each unit for each of the themes or variables in the codebook. This produces a unit-by-variable matrix that can be analyzed using a variety of statistical techniques. For example, Cowan and O'Brien (1990) tested whether males or females are more likely to be survivors in slasher films. Conventional wisdom about such films suggests that victims are mostly women and slashers are mostly men. Cowan and O'Brien selected a corpus of 56 slasher films and identified 474 victims. They coded each victim for gender and survival. They found that slashers are mostly men, but it turned out that victims are equally likely to be male or female. Women who survive are less likely to be shown engaging in sexual behavior and are less likely to be physically attractive than their nonsurviving counterparts. Male victims are cynical, egotistical, and dictatorial. Cowan and O'Brien conclude that, in slasher films, sexually pure women survive and "unmitigated masculinity" leads to death (p. 195).

The coding of texts is usually assigned to multiple coders so that the researcher can see whether the constructs being investigated are shared and whether multiple coders can reliably apply the same codes. Typically, investigators first calculate the percentage of agreement among coders for each variable or theme. They then apply a correction formula to take account of the fact that some fraction of agreement will always occur by chance. The amount of that fraction depends on the number of coders and the precision of measurement for each code. If two people code a theme present or absent, they could agree, ceteris paribus, on any answer 25% of the time by chance. If a theme, such as wealth, is measured ordinally (low, medium, high), then the likelihood of chance agreement changes accordingly. Cohen's (1960) kappa, or K, is a popular measure for taking these chances into account. When K is zero, agreement is what might be expected by chance. When K is negative, the observed level of agreement is less than one would expect by chance. How much intercoder agreement is enough? The standards are still ad hoc, but Krippendorf (1980, pp. 147-148) advocates agreement of at least .70 and notes that some scholars (e.g., Brouwer, Clark, Gerbner, & Krippendorf, 1969) use a cutoff of .80. Fleiss (1971) and Light (1971) expand kappa to handle multiple coders. For other measures of intercoder agreement, see Krippendorf (1980, pp. 147-154) and Craig (1981).

Reliability "concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials" (Carmines & Zeller, 1979, p. 11). A high level of intercoder agreement is evidence that a theme has some external validity and is not just a figment of the investigator's imagination (Mitchell, 1979). Not surprisingly, investigators have suggested many ways to assess validity (for reviews of key issues, see Campbell, 1957; Campbell & Stanley, 1963; Cook & Campbell, 1979; Denzin, 1997; Fielding & Fielding, 1986; Guba, 1981; Guba & Lincoln, 1982; Hammersley, 1992; Kirk & Miller, 1986; Lincoln & Guba, 1985). Bernard (1994) argues that, ultimately, the validity of a concept depends on the utility of the device that measures it and the collective judgment of the scientific community that a construct and its measure are valid. "In the end," he says, "we are left to deal with the effects of our judgments, which is just as it should be. Valid measurement makes valid data, but validity itself depends on the collective opinion of researchers" (p. 43). *Generalizability* refers to the degree to which the findings are applicable to other populations or samples. It draws on the degree to which the original data were representative of a larger population.

For reviews of work in content analysis, see Pool (1959); Gerbner, Holsti, Krippendorf, Paisley, and Stone (1969); Holsti (1969); Krippendorf (1980); Weber (1990); and Roberts (1997). Examples of classical content analysis can be found in media studies (Hirschman, 1987; Kolbe & Albanese, 1996; Spiggle, 1986), political rhetoric (Kaid, Tedesco, & McKinnon, 1996), folklore (Johnson & Price-Williams, 1997), business relations (Spears, Mowen, & Chakraborty, 1996), health care delivery (Potts, Runyan, Zerger, & Marchetti, 1996; Sleath, Svarstad, & Roter, 1997), and law (Imrich, Mullin, & Linz, 1995). Classical content analysis is also the fundamental means by which anthropologists test cross-cultural hypotheses (Bradley, Moore, Burton, & White, 1990; Ember & Ember, 1992; White & Burton, 1988). For early, but fundamental, criticisms of the approach, see Kracauer (1953) and George (1959).

Content Dictionaries

Computer-based, general-purpose content analysis dictionaries allow investigators to automate the coding of texts. To build such dictionaries, researchers assign words, by hand, to one or more categories (there are typically 50-60 categories in computerized content analysis dictionaries)

according to a set of rules. The rules are part of a computer program that parses new texts, assigning words to categories.

Work on content dictionaries began in the 1960s with the General Inquirer and continues to this day (Kelly & Stone, 1975; Stone et al., 1966; Zuell, Weber, & Mohler, 1989). The General Inquirer is a computer program that uses a dictionary (the *Harvard Psychosocial Dictionary*) to parse and assign text to coded categories. Over time, the dictionary has been updated. The latest version (*Harvard IV*) contains more than 10,000 words and can distinguish among multiple meanings of words (Rosenberg, Schnurr, & Oxman, 1990, p. 303). Because such dictionaries do not contain all the words in the English language, investigators can assign unrecognized words to categories as they see fit, a process of further modifying the "codebook."

How effective are computer-based dictionaries? An early version of the General Inquirer was tested on 66 suicide notes—33 written by men who had actually taken their own lives and 33 written by men who were asked to produce simulated suicide notes. The program parsed the texts and picked the actual suicide notes 91% of the time (Ogilvie, Stone, & Schneidman, 1966). Content dictionaries do not need to be very big to be useful. Colby (1966) created a simple dictionary to distinguish between Navaho and Zuni responses to thematic apperception tests. For additional examples of special-purpose dictionaries in content analysis, see Fan and Shaffer (1990), Furbee (1996), Holsti (1966), Jehn and Werner (1993), Laffal (1990, 1995), McTavish and Pirro (1990), and Schnurr, Rosenberg, Oxman, and Tucker (1986).

Content dictionaries are attractive because they are entirely reliable and automated, but, as Shapiro (1997) argues, this may be offset by a decrease in validity. For the time being, only humans can parse certain subtleties of meaning reflected in context (Viney, 1983), but computer-based dictionaries are getting better all the time. For example, texts are now scored by computer for the Gottschalk-Gleser psychological scales (measuring various forms of anxiety and hostility) with greater than .80 reliability (Gottschalk & Bechtel, 1993).

Analytic Induction and Boolean Tests

Analytic induction is a formal, nonquantitative method for building up causal explanations of phenomena from a close examination of cases. It was proposed as an alternative to statistical analysis by Znaniecki (1934,

pp. 249-331), modified by Lindesmith (1947/1968) and Cressey (1953/1971), and is discussed by Denzin (1978), Bulmer (1979), Manning (1982), and Becker (1998), among others. (For critiques of the approach, see Robinson, 1951.) The method is a formal kind of negative case analysis.

The technique can be described in a series of steps: First, define a phenomenon that requires explanation and propose an explanation. Next, examine a case to see if the explanation fits. If it does, then examine another case. An explanation is accepted until a new case falsifies it. When a case is found that doesn't fit, then, under the rules of analytic induction, the alternatives are to change the explanation (so that you can include the new case) or redefine the phenomenon (so that you exclude the nuisance case). Ideally, the process continues until a universal explanation for all known cases of a phenomenon is attained. Explaining cases by declaring them all unique is a tempting but illegitimate option. Classic examples of analytic induction include Lindesmith's (1947/1968) study of drug addicts, Cressey's (1953/1971) study of embezzlers, and McCleary's (1978) study of how parole officers decide when one of their charges is in violation of parole. For a particularly clear example of the technique, see Bloor's (1976, 1978) analysis of how doctors decide whether or not to remove children's tonsils.

Ragin (1987, 1994) formalized the logic of analytic induction, using a Boolean approach, and Romme (1995) applies the approach to textual data. Boolean algebra involves just two states (true and false, present and absent), but even with such simple inputs, things can get very complicated, very quickly. With just three dichotomous causal conditions (A and not A, B and not B, and C and not C) and one outcome variable (D and not D), there are 16 possible cases: A, B, C, D; A, not B, C, D; A, B, not C, D; and so on. Boolean analysis involves setting up what is known as a truth table, or a matrix of the actual versus the possible outcomes. (For more on truth tables and how they are related to negative case analysis, see Becker, 1998, pp. 146-214.)

Schweizer (1991, 1996) applied this method in his analysis of conflict and social status among residents of Chen Village, China. (For a discussion of Schweizer's data collection and analysis methods, see Bernard & Ryan, 1998.) All the data about the actors in this political drama were extracted from a historical narrative about Chen Village. Like classic content analysis and cognitive mapping, analytic induction requires that human coders

read and code text and then produce an event-by-variable matrix. The object of the analysis, however, is not to show the relationships among all codes, but to find the minimal set of logical relationships among the concepts that accounts for a single dependent variable. With more than three variables, the analysis becomes much more difficult. Computer programs such as QCA (Drass, 1980) and ANTHROPAC (Borgatti, 1992) test all possible multivariate hypotheses and find the optimal solution. (QCA is reviewed in Weitzman & Miles, 1995.)

Ethnographic Decision Models

Ethnographic decision models (EDMs) are qualitative, causal analyses that predict behavioral choices under specific circumstances. An EDM, often referred to as a decision tree or flowchart, comprises a series of nested *if-then* statements that link criteria (and combinations of criteria) to the behavior of interest (Figure 7.7). EDMs have been used to explain how fishermen decide where to fish (Gatewood, 1983), what prices people decide to place on their products (Gladwin, 1971; Quinn, 1978), and which treatments people choose for an illness (Mathews & Hill, 1990; Ryan & Martínez, 1996; Young, 1980).

EDMs combine many of the techniques employed in grounded theory and classic content analysis. Gladwin (1989) lays out the fundamental steps for building an ethnographic decision tree model. (For other clear descriptions of the steps, see Hill, 1998; Ryan & Martínez, 1996.)

EDMs require exploratory data collection, preliminary model building, and model testing. First, researchers identify the decisions they want to explore and the alternatives that are available. Typically, EDMs are done on simple yes/no types of behaviors. They can be used, however, to predict multiple behaviors (Mathews & Hill, 1990; Young, 1980) as well as the order of multiple behaviors (Ryan & Martínez, 1996).

Next, the researchers conduct open-ended interviews to discover the criteria people use to select among alternatives. The researchers first ask people to recall the most recent example of an actual—not a hypothetical—behavior and to recall *why* they did or did not do the behavior. Here is an example from a study we've done recently: "Think about the last time you had a can of something to drink in your hand—soda, juice, water, beer, whatever. Did you recycle the can? Why [Why not]?" This kind of question generates a list of *decision criteria*. To

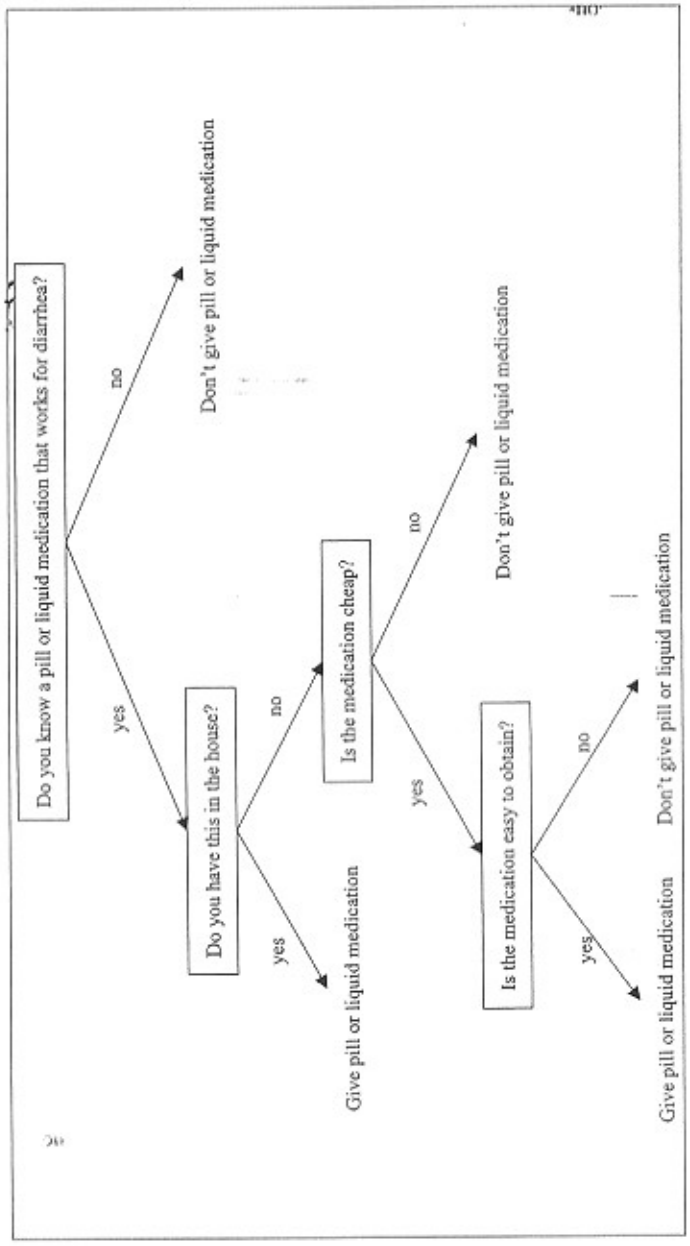

Figure 7.7. Decision Model of Constraints on the Use of Pills or Liquid Medications for Mothers Treating Children with Diarrhea in Rural Mexico
SOURCE: Based on data in Ryan and Martínez (1996).

understand how these criteria might be linked, EDM researchers ask people to compare the latest decision with other similar decisions made in the past. Some researchers have used vignettes to elicit the relationships among criteria (e.g., Weller, Ruebush, & Klein, 1997; Young, 1980).

With a list of decision criteria in hand, the researchers' next step is to systematically collect data, preferably from a new group of people, about how each criterion applies or does not apply to a recent example of the behavior. "Was a recycling bin handy?" and "Do you normally recycle cans at home?" are 2 of the 30 questions we've asked people in our study of recycling behavior. The data from this stage are used to build a preliminary model of the decision process for the behavior under scrutiny. Cases that do not fit the model are examined closely and the model is modified. Researchers tweak, or tune, the model until they achieve a satisfactory level of postdictive accuracy—understood to be at least 80% among EDM researchers. Parsimonious models are favored over more complicated ones. (For automated ways of building and pruning decision trees, see Mingers, 1989a, 1989b.)

The process doesn't end there—the same data are used in building a preliminary model and in testing its postdictive accuracy. When EDM researchers feel confident in their model, they test it on an independent sample to see if it predicts as well as it postdicts. Typically, EDMs predict more than 80% of whatever behavior is being modeled, far above what we expect by chance. (For more detailed arguments on how to calculate accuracy in EDMs, see Ryan & Martínez, 1996; Weller et al., 1997.)

Because of the intensive labor involved, EDMs have been necessarily restricted to relatively simple decisions in relatively small and homogeneous populations. Recently, however, we found we could effectively test, on a nationally representative sample, our ethnographically derived decision models for whether or not to recycle cans and whether or not to ask for paper or plastic bags at the grocery store (Bernard, Ryan, & Borgatti, 1999).

EDMs can be displayed as decision trees (e.g., Gladwin, 1989), as decision tables (Mathews & Hill, 1990; Young, 1980), or as sets of rules in the form of *if-then* statements (Ryan & Martínez, 1996). Like componential analysis, folk taxonomies, and schema analysis, EDMs represent an aggregate decision process and do not necessarily represent what is going on inside people's heads (Garro, 1998).

◆ Breaking Down the Boundaries

Text analysis as a research strategy permeates the social sciences, and the range of methods for conducting text analysis is inspiring. Investigators examine words, sentences, paragraphs, pages, documents, ideas, meanings, paralinguistic features, and even what is missing from the text. They interpret, mark, retrieve, and count. By turns, they apply interpretive analysis and numerical analysis. They use text analysis for exploratory and confirmatory purposes. Researchers identify themes, describe them, and compare them across cases and groups. Finally, they combine themes into conceptual models and theories to explain and predict social phenomena.

Figure 7.1 depicts a broad range of analysis techniques found across the social sciences. To conform our presentation with the literature on qualitative methods, we have organized these techniques according to the goals of the investigators and the kinds of texts to which the techniques are typically applied.

In this chapter, we focus on the sociological tradition that uses text as a "window into experience" rather than the linguistic tradition that describes how texts are developed and structured. Texts such as conversations, performances, and narratives are analyzed by investigators from both the sociological and linguistic traditions. Although the agendas of the investigators may differ, we see no reason why many of the sociological techniques we describe could not be useful in the linguistic tradition and vice versa.

We also distinguish between those analyses associated with systematically elicited data and those associated with free-flowing texts. We argue, however, that these data-analytic pairings are ones of convention rather than necessity. Investigators want to (a) identify the range and salience of key items and concepts, (b) discover the relationships among these items and concepts, and (c) build and test models linking these concepts together. They use free-listing tasks, KWIC, word counts, and the exploratory phases of grounded theory, schema analysis, and EDM to discover potentially useful themes and concepts.

Researchers use pile sorts, paired comparisons, triads tests, frame substitution tasks, semantic networks, cognitive maps, content analysis and content dictionaries, and the modeling phases of grounded theory, schema analysis, and EDM to discover how abstract concepts are related to each other. They display the relationships as models or frameworks. These

frameworks include formal models that rely on Boolean logic (componential analysis and analytic induction), hierarchical models (taxonomies and ethnographic decision models), probabilistic models (classic content analysis and content dictionaries), and more abstract models such as those produced by grounded theory and schema analysis. Below we describe two important examples of studies in which researchers combined methods to understand their data more fully.

Jehn and Doucet (1996, 1997) used word counts, classical content analysis, and mental mapping to examine conflicts among Chinese and U.S. business associates. They asked 76 U.S. managers who had worked in Sino-American joint ventures to describe recent interpersonal conflicts with business partners. Each person described a situation with a same-culture manager and a different-cultural manager. The researchers made sure that each manager interviewed included information about his or her relationship to the other person, who was involved, what the conflict was about, what caused the conflict, and how the conflict was resolved.

After collecting the narratives, Jehn and Doucet asked their informants to help identify the emic themes in the narratives. First, they generated separate lists of words from the intercultural and intracultural conflict narratives. They asked three expatriate managers to act as judges and to identify all the words that were related to conflict. They settled on a list of 542 conflict words from the intercultural list and 242 conflict words from the intracultural list. Jehn and Doucet then asked the three judges to sort the words into piles or categories. The experts identified 15 subcategories for the intercultural data (things like *conflict, expectations, rules, power,* and *volatile*) and 15 categories for the intracultural data (things like *conflict, needs, standards, power, contentious,* and *lose*). Taking into consideration the total number of words in each corpus, conflict words were used more in intracultural interviews and resolution terms were more likely to be used in intercultural interviews.

Jehn and Doucet also used traditional content analysis on their data. The had two coders read the 152 conflict scenarios (76 intracultural and 76 intercultural) and evaluate (on a 5-point scale) each on 27 different themes they had identified from the literature. This produced two 76 × 27 scenario-by-theme profile matrices—one for the intracultural conflicts and one for the intercultural conflicts. The first three factors from the intercultural matrix reflect (a) interpersonal animosity and hostility, (b) aggravation, and (c) the volatile nature of the conflict. The first two

factors from the intracultural matrix reflect (a) hatred and animosity with a volatile nature and (b) conflicts conducted calmly with little verbal intensity.

Finally, Jehn and Doucet identified the 30 intracultural and the 30 intercultural scenarios that they felt were the clearest and pithiest. They recruited 50 *more* expatriate managers to assess the similarities (on a 5-point scale) of 60-120 randomly selected pairs of scenarios. When combined across informants, the managers' judgments produced two aggregate, scenario-by-scenario, similarity matrices—one for the intracultural conflicts and one for the intercultural conflicts. Multidimensional scaling of the intercultural similarity data identified four dimensions: (a) open versus resistant to change, (b) situational causes versus individual traits, (c) high- versus low-resolution potential based on trust, and (d) high- versus low-resolution potential based on patience. Scaling of the intracultural similarity data identified four different dimensions: (a) high versus low cooperation, (b) high versus low confrontation, (c) problem solving versus accepting, and (d) resolved versus ongoing.

The work of Jehn and Doucet is impressive because the analysis of the data from these tasks produced different sets of themes. All three emically induced theme sets have some intuitive appeal, and all three yield analytic results that are useful. The researchers could have also used the techniques of grounded theory or schema analysis to discover even more themes.

Jehn and Doucet are not the only researchers ever to combine different analytic techniques. In a series of articles on young adult "occasional" drug users, Agar (1979, 1980, 1983) used grounded theory methods to build models of behavior. He then used classical content analysis to test his hypotheses. Agar conducted and transcribed three interviews with each of his three informants. In his 1979 article, Agar describes his initial, intuitive analysis. He pulled all the statements that pertained to informants' interactions or assessments of other people. He then looked at the statements and sorted them into piles based on their content. He named each pile as a theme and assessed how the themes interacted. He found that he had three piles. The first contained statements in which the informant was expressing negative feelings toward a person in a dominant social position. The second was made up of statements emphasizing the other's knowledge or awareness. The statements in the third small cluster emphasized the importance of change or openness to new experiences.

From this intuitive analysis, Agar felt that his informants were telling him that those in authority were only interested in displaying their authority unless they had knowledge or awareness; that knowledge or awareness comes through openness to new experience; and that most in authority are closed to new experience or change.

To test his intuitive understanding of the data, Agar (1983) used all the statements from a single informant and coded the statements for their role type (kin, friend/acquaintance, educational, occupational, or other), power (dominant, symmetrical, subordinate, or undetermined), and affect (positive, negative, ambivalent, or absent). Agar was particularly interested in whether negative sentiments were expressed toward those in dominant social roles. For one informant, Agar found that out of 40 statements coded as dominant, 32 were coded negative and 8 were coded positive. For the 36 statements coded as symmetrical, 20 were coded positive and 16 negative, lending support to his original theory.

Next, Agar looked closely at the deviant cases—the 8 statements where the informant expressed positive affect toward a person in a dominant role. These counterexamples suggested that the positive affect was expressed toward a dominant social other when the social other possessed, or was communicating to the informant, knowledge that the informant valued.

Finally, Agar (1980) developed a more systematic questionnaire to test his hypothesis further. He selected 12 statements, 4 from each of the control, knowledge, and change themes identified earlier. He matched these statements with eight roles from the informant's transcript (father, mother, employer, teacher, friend, wife, coworker, and teammate). Agar then returned to his informant and asked if the resulting statements were true, false, or irrelevant. (In no case did the informant report "irrelevant.") Agar then compared the informant's responses to his original hypotheses. He found that on balance his hypotheses were correct, but discrepancies between his expectations and his results suggested areas for further research.

These examples show that investigators can apply one technique to different kinds of data and they can apply multiple techniques to the same data set. Text analysis is used by avowed positivists and interpretivists alike. As we have argued elsewhere (Bernard, 1993; Bernard & Ryan, 1998), methods are simply tools that belong to everyone.

◆ What's Next?

We do not want to minimize the profound *intellectual* differences in the epistemological positions of positivists and interpretivists. We think, however, that when researchers can move easily and cheaply between qualitative and quantitative data collection and analysis, the distinctions between the two epistemological positions will become of less *practical* importance. That is, as researchers recognize the full array of tools at their disposal, and as these tools become easier to use, the pragmatics of research will lessen the distinction between qualitative and quantitative data and analysis.

The process is under way—and is moving fast—with the development of increasingly useful software tools for qualitative data analysis. Useful tools create markets, and market needs create increasingly useful tools. Qualitative data analysis packages (ATLAS/ti, NUD•IST, Code-A-Text, the Ethnograph, AnSWR, and others) have improved dramatically over the past few years (Fischer, 1994; Kelle, 1995; Weitzman & Miles, 1995). These products, and others, make it easier and easier for researchers to identify themes, build codebooks, mark text, create memos, and develop theoretical models. Based loosely on a grounded theory type of approach to qualitative analysis, many program suites have recently folded in techniques from classical content analysis. Several programs, for example, allow researchers to export data to matrices that they can then analyze using other programs.

Investigators, however, remain constrained by program-defined units of analysis—usually marked blocks of text or informants. Researchers need the flexibility to create matrices on demand, whether they be word-by-theme or word-by-informant matrices for word analysis and sentence-by-code or paragraph-by-code matrices for content analysis. A series of word analysis functions would greatly enhance the automated coding features found in programs that are geared to the interests of scholars in the grounded theory school. Investigators should be able to code a section of text using grounded theory, then identify the key words associated with each theme. They should be able to use key words to search for additional occurrences of the theme in large corpuses of text.

When programs make it easy to use multiple coders and to identify intercoder agreements and disagreements systematically, researchers will be better able to describe themes and to train assistants. Adding a variety of measures for calculating intercoder agreement, which only some programs do, would also be helpful. Some programs offer researchers the option of recording the marking behavior of multiple coders, yet offer no direct way to measure intercoder agreement.

The evolution of text analysis software is just beginning. Some 15 years ago, spell checkers, thesauruses, and scalable fonts were all sold separately. Today, these functions are integrated into all full-featured word-processing packages. Just 10 years ago, graphics programs were sold separately from programs that do statistical analysis. Today, graphics functions are integrated into all full-featured packages for statistical analysis. As programmers of text analysis software compete for market share, packages will become more inclusive, incorporating methods from both sides of the epistemological divide. It can't happen too soon.

◆ Notes

1. MDS displays are highly evocative. They beg to be interpreted. In fact, they *must* be interpreted. Why are some illnesses at the top of Figure 7.3 and some at the bottom? We think the illnesses at the top are more of the chronic variety, whereas those at the bottom are more acute. We also think that the illnesses on the left are less serious than those on the right. We can test ideas like these by asking key informants to help us understand the arrangement of the illnesses in the MDS plot. (For more examples of mental maps, see Albert, 1991; D'Andrade et al., 1972; Erickson, 1997.) (There is a formal method, called property fitting analysis, or PROFIT, for testing ideas about the distribution of items in an MDS map. This method is based on linear regression. See Kruskal & Wish, 1978.)

2. Alternatively, profile matrices (the usual thing-by-variable attribute matrix ubiquitous in the social sciences) can be converted to similarity matrices (thing-by-thing matrices in which the cells contain measures of similarity among pairs of things) and then analyzed with MDS (for step-by-step instructions, see Borgatti, 1999).

◆ References

Agar, M. (1979). Themes revisited: Some problems in cognitive anthropology. *Discourse Processes, 2,* 11-31.

Agar, M. (1980). Getting better quality stuff: Methodological competition in an interdisciplinary niche. *Urban Life, 9,* 34-50.

Agar, M. (1983). Microcomputers as field tools. *Computers and the Humanities, 17,* 19-26.

Agar, M. (1996). *Speaking of ethnography* (2nd ed.). Thousand Oaks, CA: Sage.