Trotter, M. (1992). *Life writing: Exploring the practice of autoethnography in anthropology.* Unpublished master's thesis, University of Illinois, Urbana-Champaign.

True confessions: The age of the literary memoir [Special issue]. (1996, May 12). *New York Times Magazine.*

Turner, V., & Bruner, E. (Eds.). (1986). *The anthropology of experience.* Urbana: University of Illinois Press.

Tyler, S. (1986). Post-modern ethnography: From document of the occult to occult document. In J. Clifford and G. E. Marcus (Eds.), *Writing culture: The poetics and politics of ethnography* (pp. 122-140). Berkeley: University of California Press.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

Van Maanen, J. (1995). An end to innocence: The ethnography of ethnography. In J. Van Maanen (Ed.), *Representation in ethnography* (pp. 1-35). Thousand Oaks, CA: Sage.

Van Maanen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy.* Albany: State University of New York Press.

Wittgenstein, L. (1953). *Philosophical investigations* (G. Anscombe, Trans.). New York: Macmillan.

Zola, I. K. (1982). *Missing pieces: A chronicle of living with a disability.* Philadelphia: Temple University Press.

Zussman, R. (1996). Autobiographical occasions. *Contemporary Sociology, 25,* 143-148.

7

Data Management and Analysis Methods

Gery W. Ryan and H. Russell Bernard

◆ Texts Are Us

This chapter is about methods for managing and analyzing qualitative data. By *qualitative data* we mean text: newspapers, movies, sitcoms, e-mail traffic, folktales, life histories. We also mean narratives—narratives about getting divorced, about being sick, about surviving hand-to-hand combat, about selling sex, about trying to quit smoking. In fact, most of the archaeologically recoverable information about human thought and human behavior is text, the "good stuff" of social science.

Scholars in content analysis began using computers in the 1950s to do statistical analysis of texts (Pool, 1959), but recent advances in technology are changing the economics of the social sciences. Optical scanning today makes light work of converting written texts to machine-readable form. Within a few years, voice-recognition software will make light work of transcribing open-ended interviews. These technologies are blind to epistemological differences. Interpretivists and positivists alike are using these technologies for the analysis of texts, and will do so more and more.

Like Tesch (1990), we distinguish between the *linguistic tradition,* which treats text as an object of analysis itself, and the *sociological tradition,* which treats text as a window into human experience (see Figure 7.1). The linguistic tradition includes narrative analysis, conversation (or

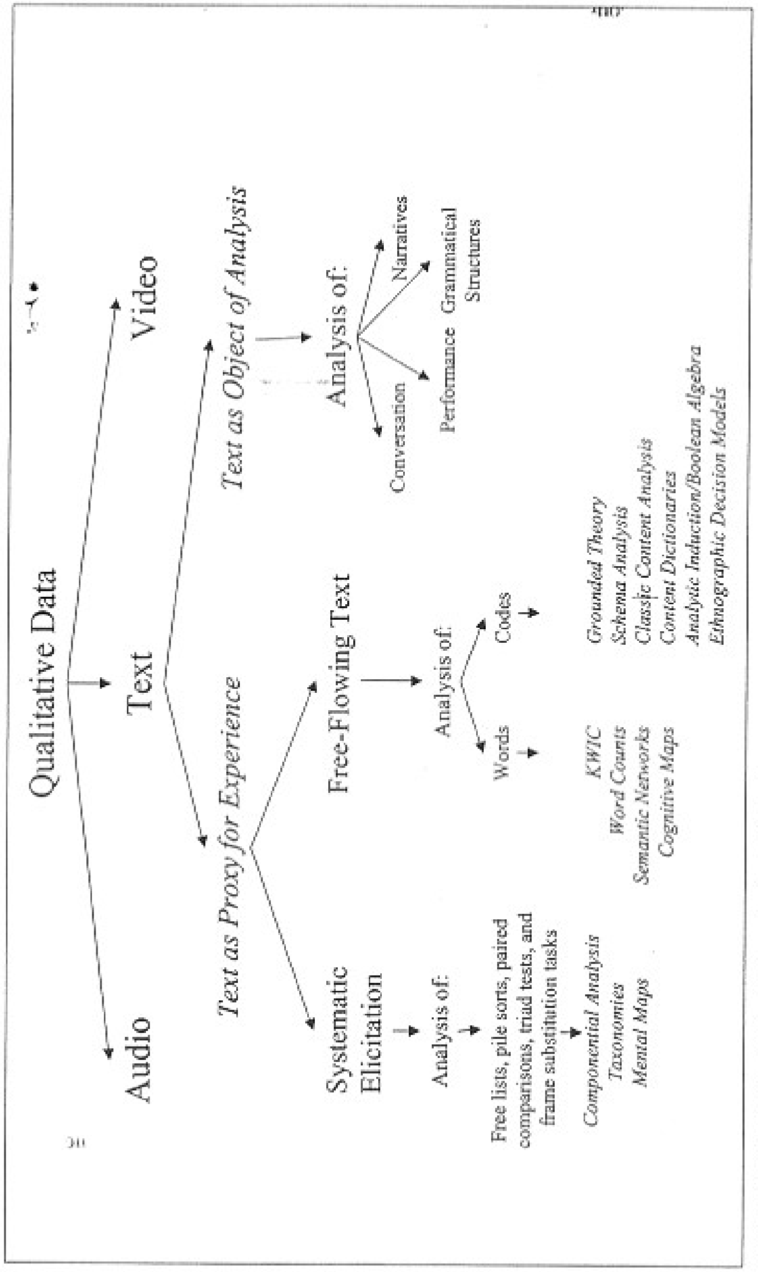

Figure 7.1. Typology of Qualitative Analysis Techniques

discourse) analysis, performance analysis, and formal linguistic analysis. Methods for analyses in this tradition are covered elsewhere in this *Handbook*. We focus here on methods used in the sociological tradition, which we take to include work across the social sciences.

There are two kinds of written texts in the sociological tradition: (a) words or phrases generated by techniques for systematic elicitation and (b) free-flowing texts, such as narratives, discourse, and responses to open-ended interview questions. In the next section, we describe some methods for collecting and analyzing words or phrases. Techniques for data collection include free lists, pile sorts, frame elicitations, and triad tests. Techniques for the analysis of these kinds of data include componential analysis, taxonomies, and mental maps.

We then turn to the analysis of free-flowing texts. We look first at methods that use raw text as their input—methods such as key-words-in-context, word counts, semantic network analysis, and cognitive maps. We then describe methods that require the reduction of text to codes. These include grounded theory, schema analysis, classical content analysis, content dictionaries, analytic induction, and ethnographic decision models. Each of these methods of analysis has advantages and disadvantages. Some are appropriate for exploring data, others for making comparisons, and others for building and testing models. Nothing does it all.

◆ Collecting and Analyzing Words or Phrases

Techniques for Systematic Elicitation

Researchers use techniques for systematic elicitation to identify lists of items that belong in a cultural domain and to assess the relationships among these items (for detailed reviews of these methods, see Bernard, 1994; Borgatti, 1998; Weller, 1998; Weller & Romney, 1988). Cultural domains comprise lists of words in a language that somehow "belong together." Some domains (such as animals, illnesses, things to eat) are very large and inclusive, whereas others (animals you can keep at home, illnesses that children get, brands of beer) are relatively small. Some lists (such as the list of terms for members of a family or the names of all the Major League Baseball teams) are agreed on by all native speakers of a language; others (such as the list of carpenters' tools) represent highly specialized knowledge, and still others (like the list of great left-handed baseball pitchers of the 20th century) are matters of heated debate. Below we review some of the most common systematic elicitation techniques and discuss how researchers analyze the data they generate.

Free Lists

Free lists are particularly useful for identifying the items in a cultural domain. To elicit domains, researchers might ask, "What kinds of illnesses do you know?" Some short, open-ended questions on surveys can be considered free lists, as can some responses generated from in-depth ethnographic interviews and focus groups. Investigators interpret the frequency of mention and the order in which items are mentioned in the lists as indicators of items' salience (for measures of salience, see Robbins & Nolan, 1997; Smith, 1993; Smith & Borgatti, 1998). The co-occurrence of items across lists and the proximity with which items appear in lists may be used as measures of similarity among items (Borgatti, 1998; Henley, 1969; for a clear example, see Fleisher & Harrington, 1998).

Paired Comparisons, Pile Sorts, Triad Tests

Researchers use paired comparisons, pile sorts, and triads tests to explore the *relationships* among items. Here are two questions we might ask someone in a paired comparison test about a list of fruits: (a) "On a scale of 1 to 5, how similar are lemons and watermelons with regard to sweetness?" (b) "Which is sweeter, watermelons or lemons?" The first question produces a set of fruit-by-fruit matrices, one for each respondent, the entries of which are scale values on the similarity of sweetness among all pairs of fruits. The second question produces, for each respondent, a perfect rank ordering of the set of fruits.

In a pile sort, the researcher asks each respondent to sort a set of cards or objects into piles. Item similarity is the number of times each pair of items is placed in the same pile (for examples, see Boster, 1994; Roos, 1998). In a triad test, the researcher presents sets of three items and asks each respondent either to "choose the two most similar items" or to "pick the item that is the most different." The similarity among pairs of items is the number of times people choose to keep pairs of items together (for some good examples, see Albert, 1991; Harman, 1998).

Frame Substitution

In the frame substitution task (D'Andrade, 1995; D'Andrade, Quinn, Nerlove, & Romney, 1972; Frake, 1964; Metzger & Williams, 1966), the researcher asks the respondent to link each item in a list of items with a list of attributes. D'Andrade et al. (1972) gave people a list of 30 illness terms and asked them to fill in the blanks in frames such as "You can catch _____ from other people," "You can have _____ and never know it," and "Most people get _____ at one time or other" (p. 12; for other examples of frame substitution, see Furbee & Benfer, 1983; Young, 1978).

Techniques for Analyzing Data About Cultural Domains

Researchers use these kinds of data to build several kinds of models about how people think. *Componential analysis* produces formal models of the elements in a cultural domain, and *taxonomies* display hierarchical associations among the elements in a domain. *Mental maps* are best for displaying fuzzy constructs and dimensions. We treat these in turn.

Componential Analysis

As we have outlined elsewhere, componential analysis (or feature analysis) is a formal, qualitative technique for studying the content of meaning (Bernard, 1994; Bernard & Ryan, 1998). Developed by linguists to identify the features and rules that distinguish one sound from another (Jakobson & Halle, 1956), the technique was elaborated by anthropologists in the 1950s and 1960s (Conklin, 1955; D'Andrade, 1995; Frake, 1962; Goodenough, 1956; Rushforth, 1982; Wallace, 1962). (For a particularly good description of how to apply the method, see Spradley, 1979, pp. 173-184.)

Componential analysis is based on the principle of distinctive features. Any two items (sounds, kinship terms, names of plants, names of animals, and so on) can be distinguished by some minimal set ($2n$) of binary features—that is, features that either occur or do not occur. It takes two features to distinguish four items ($2^2 = 4$, in other words), three features to distinguish eight items ($2^3 = 8$), and so on. The trick is to identify the smallest set of features that best describes the domain of interest. Table 7.1 shows that just three features are needed to describe kinds of horses.

TABLE 7.1 A Componential Analysis of Six Kinds of Horses

Name	*Female*	*Neuter*	*Adult*
Mare	+	–	+
Stallion	–	–	+
Gelding	–	+	+
Foal	–	+	–
Filly	+	–	–
Colt	–	–	–

SOURCE: Adapted from D'Andrade (1995).

Componential analysis produces models based on logical relationships among features. The models do not account for variations in the meanings of terms across individuals. For example, when we tried to do a componential analysis on the terms for cattle (*bull, cow, heifer, calf, steer,* and *ox*), we found that native speakers of English in the United States (even farmers) disagreed about the differences between *cow* and *heifer,* and between *steer* and *ox.* When the relationships among items are less well defined, taxonomies or mental models may be useful. Nor is there any intimation that componential analyses reflect how "people really think."

Taxonomies

Folk taxonomies are meant to capture the hierarchical structure in sets of terms and are commonly displayed as branching tree diagrams. Figure 7.1 presents a taxonomy of our own understanding of qualitative analysis techniques. Figure 7.2 depicts a taxonomy we have adapted from Pamela Erickson's (1997) study of the perceptions among clinicians and adolescents of methods of contraception. Researchers can elicit folk taxonomies directly by using successive pile sorts (Boster, 1994; Perchonock & Werner, 1969). This involves asking people to continually subdivide the piles of a free pile sort until each item is in its own individual pile. Taxonomic models can also be created with cluster analysis on the similarity data from paired comparisons, pile sorts, and triad tests. *Hierarchical cluster analysis* (Johnson, 1967) builds a taxonomic tree where each item appears in only one group.

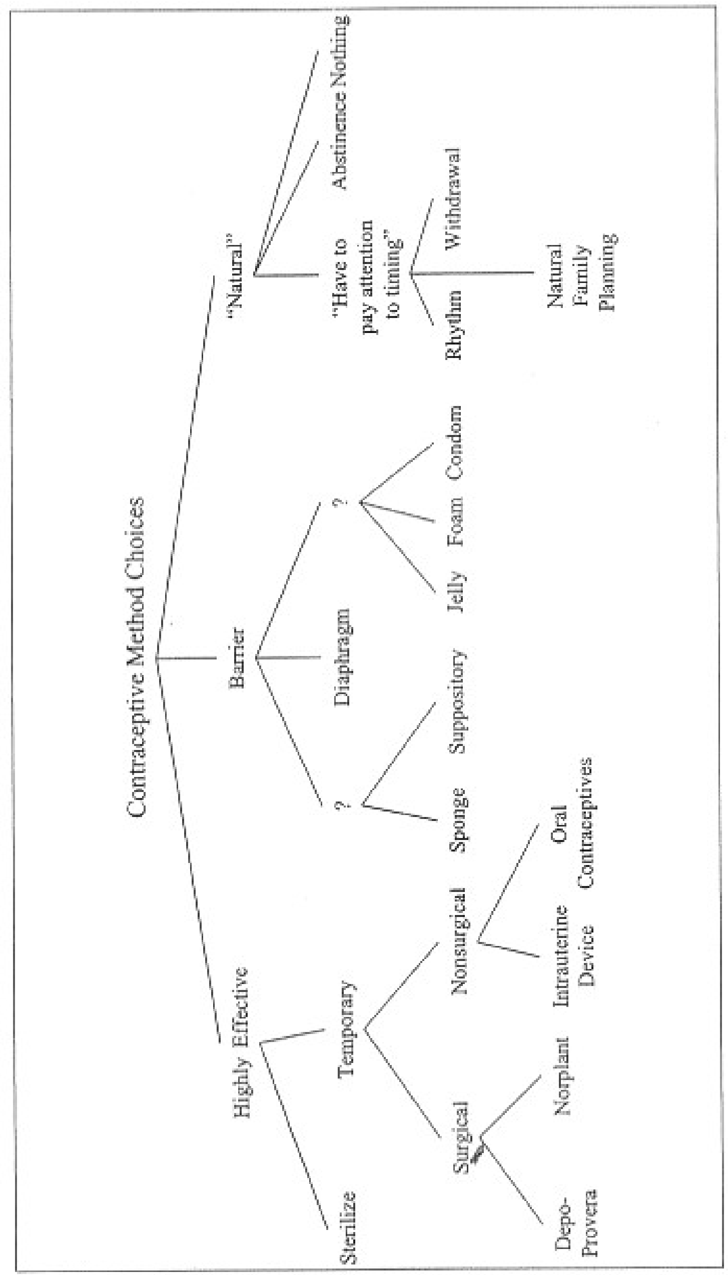

Figure 7.2. Clinicians' Taxonomy of Contraceptive Methods
SOURCE: Based on Erickson (1997).

Interinformant variation is common in folk taxonomies. That is, different people may use different words to refer to the same category of things. Some of Erickson's (1997) clinician informants referred to the "highly effective" group of methods as "safe," "more reliable," and "sure bets." Category labels need not be simple words, but may be complex phrases; for example, see the category in Figure 7.2 comprising contraceptive methods in which you "have to pay attention to timing." Sometimes, people have no labels at all for particular categories—at least none that they can dredge up easily—and categories, even when named, may be fuzzy and may overlap with other categories. *Overlapping cluster analysis* (Hartigan, 1975) identifies groups of items where a single item may appear in multiple groups.

Mental Maps

Mental maps are visual displays of the similarities among items, whether or not those items are organized hierarchically. One popular method for making these maps is by collecting data about the cognitive similarity or dissimilarity among a set of objects and then applying multidimensional scaling, or MDS, to the similarities (Kruskal & Wish, 1978).

Cognitive maps are meant to be directly analogous to physical maps. Consider a table of distances between all pairs of cities on a map. Objects (cities) that are very dissimilar have high mileage between them and are placed far apart on the map; objects that are less dissimilar have low mileage between them and are placed closer together. Pile sorts, triad tests, and paired comparison tests are measures of cognitive distance. For example, Ryan (1995) asked 11 literate Kom speakers in Cameroon to perform successive pile sorts on Kom illness terms. Figure 7.3 presents an MDS plot of the collective mental map of these terms. The five major illness categories, circled, were identified by hierarchical cluster analysis of the same matrix used to produce the MDS plot.[1]

Data from frame substitution tasks can be displayed with correspondence analysis (Weller & Romney, 1990).[2] Correspondence analysis scales both the rows and the columns into the same space. For example, Kirchler (1992) analyzed 562 obituaries of managers who had died in 1974, 1980, and 1986. He identified 31 descriptive categories from adjectives used in the obituaries and then used correspondence analysis to display how these categories were associated with men and women managers over time. Figure 7.4 shows that male managers who died in 1974 and 1980 were seen

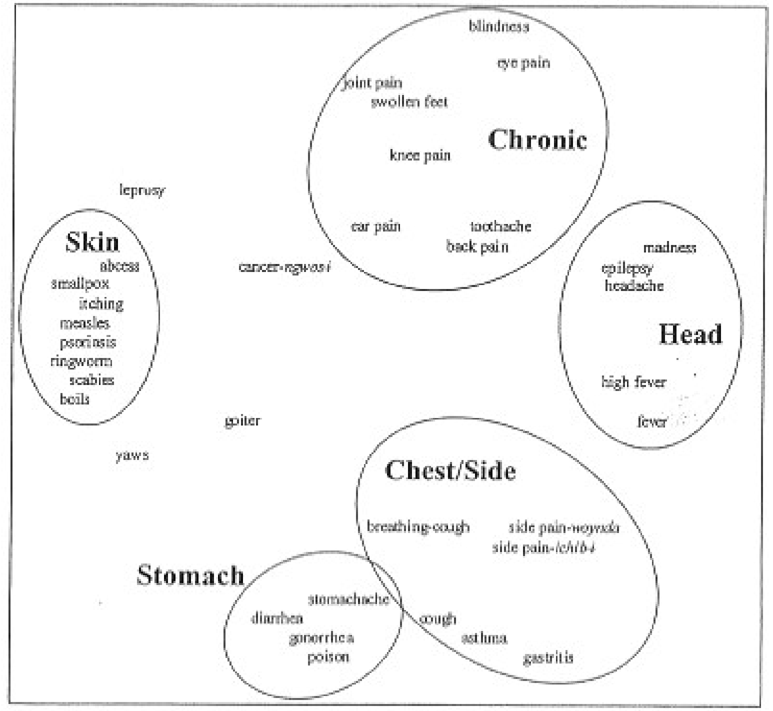

Figure 7.3. Mental Map of Kom Illness Terms

by their surviving friends and family as *active, intelligent, outstanding, conscientious,* and *experienced experts.* Although the managers who died in 1986 were still respected, they were more likely to be described as *entrepreneurs, opinion leaders,* and *decision makers.* Perceptions of female managers also changed, but they did not become more like their male counterparts. In 1974 and 1980, female managers were remembered for being nice people. They were described as *kind, likable,* and *adorable.* By 1986, women were remembered for their *courage* and *commitment.* Kirchler interpreted these data to mean that gender stereotypes changed in the early 1980s. By 1986, both male and female managers were perceived as working for success, but men impressed their colleagues through their knowledge and expertise, whereas women impressed their colleagues with motivation and engagement.

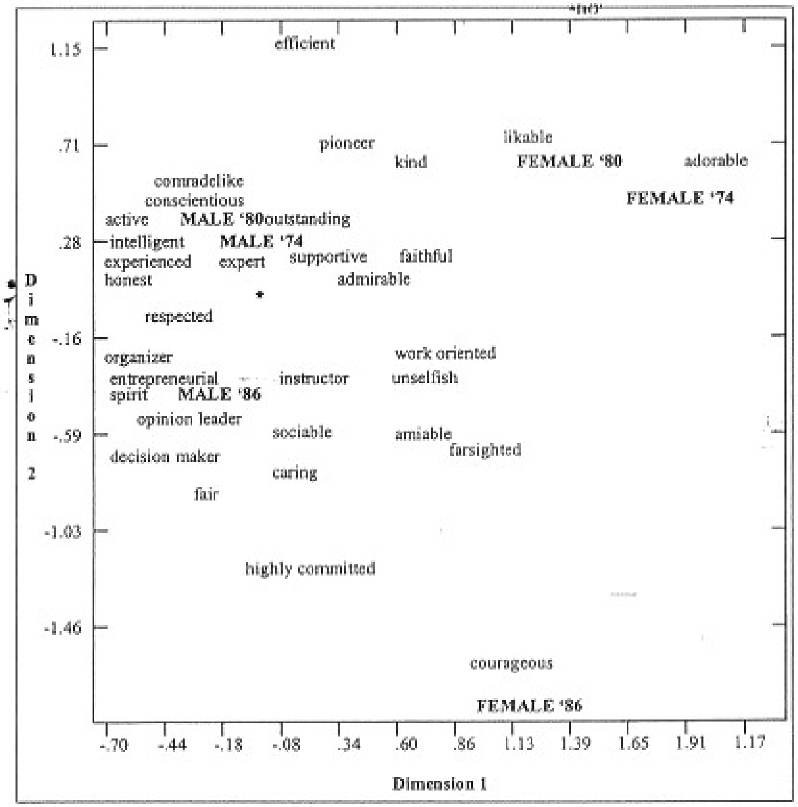

Figure 7.4. Correspondence Analysis of the Frequencies of 31 Disruptive Obituary Categories by Gender and Year of Publication

SOURCE: Erich Kirchler, "Adorable Woman, Expert Man: Changing Gender Images of Women and Men in Management," *European Journal of Social Psychology, 22* (1992), p. 371. Copyright 1992 by John Wiley & Sons Limited. Reproduced by permission of John Wiley & Sons Limited.

◆ Methods for Analyzing Free-Flowing Text

Although taxonomies, MDS maps, and the like are useful for analyzing short phrases or words, most qualitative data come in the form of free-flowing texts. There are two major types of analysis. In one, the text is segmented into its most basic meaningful components: words. In the other, meanings are found in large blocks of text.

Analyzing Words

Techniques for word analysis include key-words-in-context, word counts, structural analysis, and cognitive maps. We review each below.

Key-Words-in-Context

Researchers create key-words-in-context (KWIC) lists by finding all the places in a text where a particular word or phrase appears and printing it out in the context of some number of words (say, 30) before and after it. This produces a *concordance*. Well-known concordances have been done on sacred texts, such as the Old and New Testaments (Darton, 1976; Hatch & Redpath, 1954) and the Koran (Kassis, 1983), and on famous works of literature from Euripides (Allen & Italie, 1954) to Homer (Prendergast, 1971), to Beowulf (Bessinger, 1969), to Dylan Thomas (Farringdon & Farringdon, 1980). (On the use of concordances in modern literary studies, see Burton, 1981a, 1981b, 1982; McKinnon, 1993.)

Word Counts

Word counts are useful for discovering patterns of ideas in any body of text, from field notes to responses to open-ended questions. Students of mass media have used use word counts to trace the ebb and flow of support for political figures over time (Danielson & Lasorsa, 1997; Pool, 1952). Differences in the use of words common to the writings of James Madison and Alexander Hamilton led Mosteller and Wallace (1964) to conclude that Madison and not Hamilton had written 12 of the *Federalist Papers*. (For other examples of authorship studies, see Martindale & McKenzie, 1995; Yule 1944/1968.)

Word analysis (like constant comparison, memoing, and other techniques) can help researchers to discover themes in texts. Ryan and Weisner (1996) instructed fathers and mothers of adolescents in Los Angeles: "Describe your children. In your own words, just tell us about them." Ryan and Weisner identified all the unique words in the answers they got to that grand-tour question and noted the number of times each word was used by mothers and by fathers. Mothers, for example, were more likely to use words like *friends, creative, time,* and *honest;* fathers were more likely to use words like *school, good, lack, student, enjoys, independent,* and *extremely.* This suggests that mothers, on first mention, express concern over interpersonal issues, whereas fathers appear to prioritize achievement-oriented and individualistic issues. This kind of analysis considers neither the contexts in which the words occur nor whether the words are used negatively or positively, but distillations like these can help researchers to identify important constructs and can provide data for systematic comparisons across groups.

Structural Analysis and Semantic Networks

Network, or structural, analysis examines the properties that emerge from relations among things. As early as 1959, Charles Osgood created word co-occurrence matrices and applied factor analysis and dimensional plotting to describe the relations among words. Today, semantic network analysis is a growing field (Barnett & Danowski, 1992; Danowski, 1982, 1993). For example, Nolan and Ryan (1999) asked 59 undergraduates (30 women and 29 men) to describe their "most memorable horror film." The researchers identified the 45 most common adjectives, verbs, and nouns used across the descriptions of the films. They produced a 45(word)-by-59(person) matrix, the cells of which indicated whether each student had used each key word in his or her description. Finally, Nolan and Ryan created a 59(person)-by-59(person) similarity matrix of *people* based on the co-occurrence of the words in their descriptions.

Figure 7.5 shows the MDS of Nolan and Ryan's data. Although there is some overlap, it is pretty clear that the men and women in their study used different sets of words to describe horror films. Men were more likely to use words such as *teenager, disturbing, violence, rural, dark, country,* and *hillbilly,* whereas women were more likely to use words such as *boy, little, devil, young, horror, father,* and *evil.* Nolan and Ryan interpreted these results to mean that the men had a fear of rural people and places, whereas the women were more afraid of betrayed intimacy and spiritual possession. (For other examples of the use of word-by-word matrices, see Jang & Barnett, 1994; Schnegg & Bernard, 1996.) This example makes abundantly clear the value of turning qualitative data into quantitative data: Doing so can produce information that engenders deeper interpretations of the meanings in the original corpus of qualitative data. Just as in any mass of numbers, it is hard to see patterns in words unless one first does some kind of data reduction. More about this below.

As in word analysis, one appeal of semantic network analysis is that the data processing is done by computer. The only investigator bias intro-

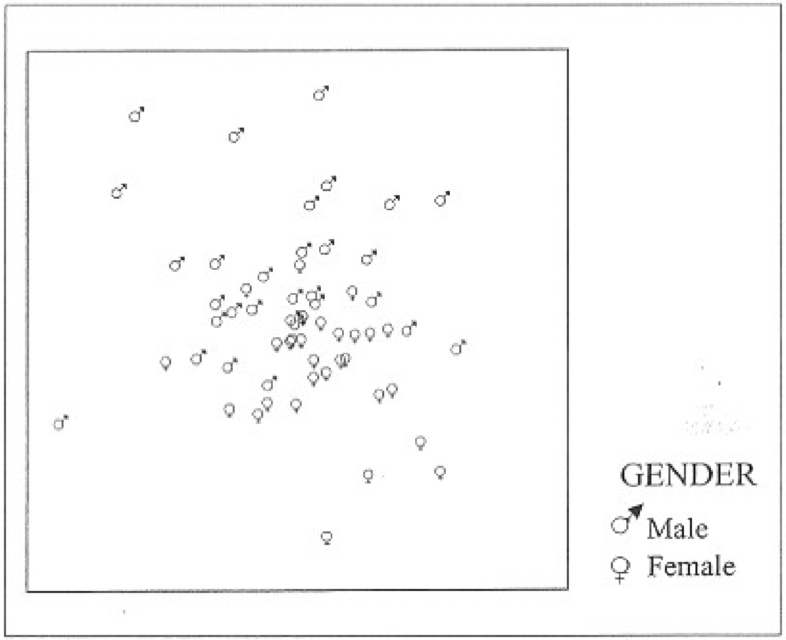

Figure 7.5. Multidimensional Scaling of Informants Based on Words Used in Descriptions of Horror Films

duced in the process is the decision to include words that occur at least 10 times or 5 times or whatever. (For discussion of computer programs that produce word-by-text and word-by-word co-occurrence matrices, see Borgatti, 1992; Doerfel & Barnett, 1996.) There is, however, no guarantee that the output of any word co-occurrence matrix will be meaningful, and it is notoriously easy to read patterns (and thus meanings) into any set of items.

Cognitive Maps

Cognitive map analysis combines the intuition of human coders with the quantitative methods of network analysis. Carley's work with this technique is instructive. Carley argues that if cognitive models or schemata exist, they are expressed in the texts of people's speech and can be represented as networks of concepts (see Carley & Palmquist, 1992,

p. 602), an approach also suggested by D'Andrade (1991). To the extent that cognitive models are widely shared, Carley asserts, even a very small set of texts will contain the information required for describing the models, especially for narrowly defined arenas of life.

In one study, Carley (1993) asked students some questions about the work of scientists. Here are two examples she collected:

> *Student A:* I found that scientists engage in research in order to make discoveries and generate new ideas. Such research by scientists is hard work and often involves collaboration with other scientists which leads to discoveries which make the scientists famous. Such collaboration may be informal, such as when they share new ideas over lunch, or formal, such as when they are coauthors of a paper.
>
> *Student B:* It was hard work to research famous scientists engaged in collaboration and I made many informal discoveries. My research showed that scientists engaged in collaboration with other scientists are coauthors of at least one paper containing their new ideas. Some scientists make formal discoveries and have new ideas. (p. 89)

Carley compared the students' texts by analyzing 11 concepts: *I, scientists, research, hard work, collaboration, discoveries, new ideas, formal, informal, coauthors, paper.* She coded the concepts for their strength, sign (positive or negative), and direction (whether one concept is logically prior to others), not just for their existence. She found that although students used the same concepts in their texts, the concepts clearly had different meanings. To display the differences in understandings, Carley advocates the use of maps that show the relations between and among concepts. Figure 7.6 shows Carley's maps of two of the texts.

Carley's approach is promising because it combines the automation of word counts with the sensitivity of human intuition and interpretation. As Carley recognizes, however, a lot depends on who does the coding. Different coders will produce different maps by making different coding choices. In the end, native-language competence is one of the fundamental methodological requirements for analysis (see also Carley, 1997; Carley & Kaufer, 1993; Carley & Palmquist, 1992; Palmquist, Carley, & Dale, 1997).

Key-words-in-context, word counts, structural analysis, and cognitive maps all reduce text to the fundamental meanings of specific words. These reductions make it easy for researchers to identify general patterns and

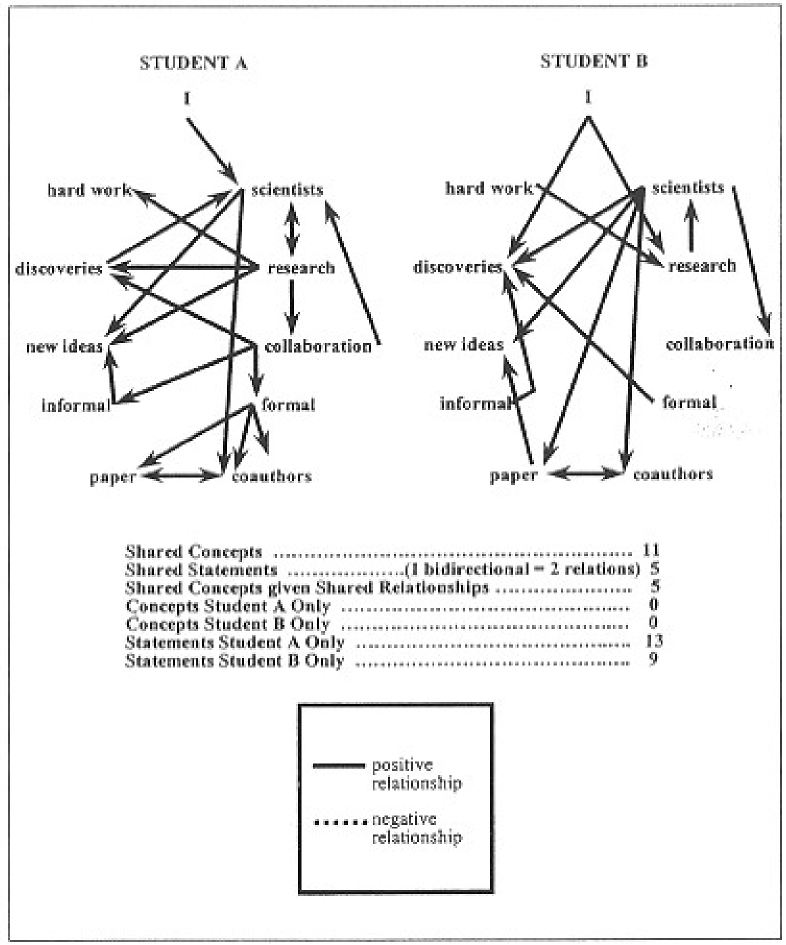

Figure 7.6. Coded Maps of Two Students' Texts

SOURCE: Kathleen Carley, "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis," in P. Marsden (Ed.), *Sociological Methodology* (Oxford: Blackwell, 1993), p. 104. Copyright 1993 by the American Sociological Association. Reproduced by permission of the American Sociological Association.

make comparisons across texts. With the exception of KWIC, however, these techniques remove words from the contexts in which they occur. Subtle nuances are likely to be lost—which brings us to the analysis of whole texts.

Analyzing Chunks of Text: Coding

Coding is the heart and soul of *whole-text analysis.* Coding forces the researcher to make judgments about the meanings of contiguous blocks of text. The fundamental tasks associated with coding are sampling, identifying themes, building codebooks, marking texts, constructing models (relationships among codes), and testing these models against empirical data. We outline each task below. We then describe some of the major coding traditions: grounded theory, schema analysis, classic content analysis, content dictionaries, analytic induction, and ethnographic decision trees. We want to emphasize that no particular tradition, whether humanistic or positivistic, has a monopoly on text analysis.

Sampling

Investigators must first identify a *corpus of texts,* and then select the *units of analysis within the texts.* Selection can be either random or purposive, but the choice is not a matter of cleaving to one epistemological tradition or another. Waitzkin and Britt (1993) did a thoroughgoing interpretive analysis of encounters between patients and doctors by selecting 50 texts at random from 336 audiotaped encounters. Trost (1986) used classical content analysis to test how the relationships between teenagers and their families might be affected by five different dichotomous variables. He intentionally selected five cases from each of the 32 possible combinations of the five variables and conducted $32 \times 5 = 160$ interviews.

Samples may also be based on extreme or deviant cases, cases that illustrate maximum variety on variables, cases that are somehow typical of a phenomenon, or cases that confirm or disconfirm a hypothesis. (For reviews of nonrandom sampling strategies, see Patton, 1990, pp. 169-186; Sandelowski, 1995b.) A single case may be sufficient to display something of substantive importance, but Morse (1994) suggests using at least six participants in studies where one is trying to understand the essence of experience. Morse also suggests 30-50 interviews for ethnographies and grounded theory studies. Finding themes and building theory may require fewer cases than comparing across groups and testing hypotheses or models.

Once the researcher has established a sample of texts, the next step is to identify the basic units of analysis. The units may be entire texts (books, interviews, responses to an open-ended question on a survey), grammatical segments (words, word senses, sentences, themes, paragraphs), formatting units (rows, columns, or pages), or simply chunks of text that reflect a single theme—what Krippendorf (1980, p. 62) calls *thematic units.* In general, where the objective is to compare across texts (as in the case of classical content analysis), the units of analysis need to be nonoverlapping. (For discussion of additional kinds of units of analysis, see Krippendorf, 1980, pp. 57-64; Tesch, 1990.)

Finding Themes

Themes are abstract (and often fuzzy) constructs that investigators identify before, during, and after data collection. Literature reviews are rich sources for themes, as are investigators' own experiences with subject matter. More often than not, however, researchers induce themes from the text itself.

There is more than one way to induce themes. Grounded theorists suggest a careful, line-by-line reading of the text while looking for processes, actions, assumptions, and consequences. Schema analysts suggest looking for metaphors, for repetitions of words, and for shifts in content (Agar & Hobbs, 1985). Content analysts have used KWIC to identify different meanings. Spradley (1979, pp. 199-201) suggests looking for evidence of social conflict, cultural contradictions, informal methods of social control, things that people do in managing impersonal social relationships, methods by which people acquire and maintain achieved and ascribed status, and information about how people solve problems. Each of these arenas is likely to yield major themes in cultures. Barkin, Ryan, and Gelberg (1999) had multiple coders independently sort informants' statements into thematic piles. They then used multidimensional scaling and cluster analysis on the pile-sort data to identify subthemes shared across coders. (For another example, see Patterson, Bettini, & Nussbaum, 1993.)

Willms et al. (1990) and Miles and Huberman (1994) suggest that researchers start with some general themes derived from reading the literature and add more themes and subthemes as they go. Shelley (1992) followed this advice in her study of how social networks affect people with end-stage kidney disease. She used the *Outline of Cultural Materials* (Murdock, 1971) as the basis of her coding scheme and then added additional themes based on a close reading of the text. Bulmer (1979) lists 10 different sources of themes, including literature reviews, professional definitions, local commonsense constructs, and researchers' values and prior experiences. He also notes that investigators' general theoretical orientations, the richness of the existing literature, and the characteristics of the phenomena being studied influence the themes researchers are likely to find.

No matter how the researcher actually *does* inductive coding, by the time he or she has identified the themes and refined them to the point where they can be applied to an entire corpus of texts, a lot of interpretive analysis has already been done. Miles and Huberman (1994) say simply, "Coding is analysis" (p. 56).

Building Codebooks

Codebooks are simply organized lists of codes (often in hierarchies). How a researcher can develop a codebook is covered in detail by Dey (1993, pp. 95-151), Crabtree and Miller (1992), and Miles and Huberman (1994, pp. 55-72). MacQueen, McLellan, Kay, and Milstein (1998) suggest that a good codebook should include a detailed description of each code, inclusion and exclusion criteria, and exemplars of real text for each theme. If a theme is particularly abstract, we suggest that the researcher also provide examples of the theme's boundaries and even some cases that are closely related but *not* included within the theme. Coding is supposed to be data *reduction*, not proliferation (Miles, 1979, pp. 593-594). The codes themselves are mnemonic devices used to identify or mark the specific themes in a text. They can be either words or numbers—whatever the researcher finds easiest to remember and to apply.

Qualitative researchers working as a team need to agree up front on what to include in their codebook. Morse (1994) suggests beginning the process with a group meeting. MacQueen et al. (1998) suggest that a single team member should be designated "Keeper of the Codebook"—we strongly agree.

Good codebooks are developed and refined as the research goes along. Kurasaki (1997) interviewed 20 *sansei*—third-generation Japanese Americans—and used a grounded theory approach to do her analysis of ethnic identity. She started with seven major themes. As the analysis progressed, she split the major themes into subthemes. Eventually, she combined two of the major themes and wound up with six major themes and a total of 18 subthemes. (Richards & Richards, 1991, discuss the theoretical principles related to hierarchical coding structures that emerge out of the data.

Araujo, 1995, uses an example from his own research on the traditional British manufacturing industry to describe the process of designing and refining hierarchical codes.)

The development and refinement of coding categories have long been central tasks in classical content analysis (see Berelson, 1952, pp. 147-168; Holsti, 1969, pp. 95-126) and are particularly important in the construction of concept dictionaries (Deese, 1969; Stone, Dunphy, Smith, & Ogilvie, 1966, pp. 134-168). Krippendorf (1980, pp. 71-84) and Carey, Morgan, and Oxtoby (1996) note that much of codebook refinement comes during the training of coders to mark the text and in the act of checking for intercoder agreement. Disagreement among multiple coders shows when the codebook is ambiguous and confusing. The first run also allows the researcher to identify good examples to include in the codebook.

Marking Texts

The act of coding involves the assigning of codes to contiguous units of text. Coding serves two distinct purposes in qualitative analysis. First, codes act as *tags* to mark off text in a corpus for later retrieval or indexing. Tags are not associated with any fixed units of text; they can mark simple phrases or extend across multiple pages. Second, codes act as *values* assigned to fixed units (see Bernard, 1991, 1994; Seidel & Kelle, 1995). Here, codes are nominal, ordinal, or ratio scale values that are applied to fixed, nonoverlapping units of analysis. The nonoverlapping units can be texts (such as paragraphs, pages, documents), episodes, cases, or persons. *Codes as tags* are associated with grounded theory and schema analysis (reviewed below). *Codes as values* are associated with classic content analysis and content dictionaries. The two types of codes are not mutually exclusive, but the use of one gloss—*code*—for both concepts can be misleading.

Analyzing Chunks of Texts: Building Conceptual Models

Once the researcher identifies a set of things (themes, concepts, beliefs, behaviors), the next step is to identify how these things are linked to each other in a theoretical model (Miles & Huberman, 1994, pp. 134-137). Models are sets of abstract constructs and the relationships among them