7

Real and Virtual Relationships in Qualitative Data Analysis

with Barbara Miller

In this chapter, we apply a realist perspective to qualitative data analysis. We do this by distinguishing two types of analytic strategies, which we label *categorizing* and *connecting* strategies. This distinction is based on the distinction between similarity and contiguity, described in Chapter 4. To recapitulate, similarity and contiguity refer to two fundamentally different kinds of relationships between things, neither of which can be assimilated to the other. Similarity relationships are based on resemblances or commonalities independent of proximity in time or space. Ontologically, similarity relationships are *virtual* relationships, based on comparison rather than actual connection or influence. Relationships of contiguity, on the other hand, presume, explicitly or implicitly, a real connection or association that is *not* a matter of similarity.

To our knowledge, the extensively developed theoretical analysis of these two types of relationships, primarily in linguistics, has never been explicitly applied to qualitative data analysis. Although the role of similarity in categorizing is often acknowledged,[1] the importance of contiguity relations in other types of analysis is rarely stated, and the similarity/contiguity distinction itself, though sometimes implicitly recognized and described, is not linked to existing theoretical work on this distinction.

---

[1] Categorization in qualitative analysis is almost always based on similarity, despite the existence of theories of categorization (e.g., Lakoff, 1987) that include contiguity-based relationships (e.g., metonymy) as well as similarity-based ones.

In terms of the concrete processes of data analysis, the difference between these two strategies was well described by Smith (1979):

> I usually start . . . at the beginning of the notes. I read along and seem to engage in two kinds of processes—comparing and contrasting, and looking for antecedents and consequences. . .
>
>   The essence of concept formation [the first process] is . . . "How are they alike, and how are they different?" The similar things are grouped and given a label that highlights their similarity. . . . In time, these similarities and differences come to represent clusters of concepts, which then organize themselves into more abstract categories and eventually into hierarchical taxonomies.
>
>   Concurrently, a related but different process is occurring. . . . The conscious search for the consequences of social items . . . seemed to flesh out a complex systemic view and a concern for process, the flow of events over time. In addition it seemed to argue for a more holistic, systemic, interdependent network of events at the concrete level and concepts and propositions at an abstract level. . . . At a practical level, while in the field, the thinking, searching, and note recording reflected not only a consciousness of similarities and differences but also an attempt to look for unexpected relationships, antecedents, and consequences within the flow of items. (p. 338)

A similar distinction is found in many accounts of qualitative data analysis. For example, Seidman (1998, p. 101 ff.) described two main strategies in the analysis of interviews: the categorization of interview material through coding and thematic analysis, and the creation of several different types of narratives, which he called "profiles" and "vignettes." Other versions of this distinction are Weiss's (1994) contrast between "issue-focused" and "case-focused" analysis, and Coffey and Atkinson's (1996) between "concepts and coding" and "narratives and stories."

However, none of these authors examined the principles on which these distinctions are based, and the similarity/contiguity distinction is frequently confounded with others. For example, Ezzy (2002, p. 95) distinguished narrative analysis from coding primarily in terms of its being more holistic, interpretive, and "in process," and as employing a constructivist approach and "situated relativity."

Jakobson's application of the similarity/contiguity distinction to aphasia (1956), discussed later in this chapter, suggests that there may be a neurological basis for this distinction. Research on memory by the psychologist Tulving (1983; Tulving & Craik, 2000) and others provides some support for this view. Tulving identified two distinct, though interacting, systems of memory, which he called semantic memory and episodic memory. Semantic memory is memory of facts, concepts, principles, and other sorts of information, organized conceptually rather than in terms of the context in which they were learned. Episodic memory, in contrast, is memory of events and episodes, organized temporally in terms of the context of their occurrence. Extensive experimental research, in both humans and other animals (Dere,

Easton, Nadel, & Huston, 2008; Shettleworth, 2010, pp. 249–256; Tulving & Craik, 2000), has supported this distinction, particularly the idea of episodic memory as a distinct system, as have studies demonstrating that memory loss may selectively affect one or the other of these systems (Mayes, 2000), and brain imaging has shown that encoding or retrieving information from the two memory systems engages different areas of the brain (Nyberg & Cabeza, 2000; Wheeler, 2000).

We begin by separately describing categorizing and connecting strategies, presenting the strengths and limitations of each approach, and applying this distinction to data displays (Miles & Huberman, 1994). We then discuss ways of integrating these two approaches. We conclude with some observations on the use of computers in qualitative data analysis.

Categorizing Strategies

The most widely used categorizing strategy in qualitative data analysis is coding. In coding, the data are segmented into discrete units (Strauss, 1987, p. 29, referred to this as "fracturing" the data), and the segments are labeled and grouped by category; they are then examined and compared, both within and between categories. Coding categories "are a means of sorting the descriptive data you have collected . . . so that the material bearing on a given topic can be physically separated from other data" (Bogdan & Biklen, 2003, p. 161). Coding and then sorting by code creates a similarity-based ordering of data that replaces the original contiguity-based ordering.

Example 7.1

Types of Categories

An important distinction among types of categories, one that also draws on a realist approach to qualitative data analysis (although not employing the similarity/contiguity distinction), is that between organizational, substantive, and theoretical categories (Maxwell, 2005, pp. 97–98). I'm not presenting these as absolute distinctions; many actual coding categories can be seen as involving aspects of more than one type, or as being intermediate between two types. However, I believe that the conceptual typology is valuable.

Organizational categories are broad areas or issues that are often established prior to data collection. McMillan and Schumacher (2001)

(Continued)

(Continued)

refer to these as *topics* rather than categories, stating that "a topic is the descriptive name for the subject matter of the segment. You are not, at this time, asking 'What is said?' which identifies the meaning of the segment" (p. 469). In a study of elementary school principals' practices of retaining children in a grade, examples of such categories are "retention," "policy," "goals," "alternatives," "and "consequences" (p. 470). Organizational categories function primarily as abstract "bins" for sorting the data for further analysis; they do not specifically address what is actually happening or what meaning these topics have for participants. They are often useful as organizational tools in your analysis, as chapter or section headings in presenting your results, but they don't by themselves provide much insight into what is actually going on (cf. Coffey & Atkinson, 1996, pp. 34–35).

This latter task requires substantive and/or theoretical categories, ones that incorporate what's actually taking place, or the actual understandings of this that participants have. These latter categories can often be seen as subcategories of the organizational ones, but they are generally *not* subcategories that, in advance, you could have known would be significant, unless you are already fairly familiar with the kind of participants or setting you're studying or are using a well-developed theory. They implicitly make some sort of claim about the phenomena being studied—that is, they could be *wrong*, rather than simply being conceptual boxes for holding data.

Substantive categories are primarily *descriptive*, in a broad sense that includes description of participants' concepts and beliefs; they stay close to the data categorized, and don't inherently imply a more abstract theory. In the study of grade retention mentioned earlier, examples of substantive categories derived from interviews with principals would be "retention as failure," "retention as a last resort," "self-confidence as a goal," "parent's willingness to try alternatives," and "not being in control (of the decision)" (McMillan & Schumacher, 2001, p. 472). Categories taken from participants' own words and concepts (what are generally called "emic" categories) are usually substantive, but many substantive categories are not emic, being based on the *researcher's* description of what's going on. Substantive categories are often inductively generated through a close "open coding" of the data (Strauss & Corbin, 1990). They can be used in *developing* a more general theory of what's going on, but they don't *depend on* this theory.

(Continued)

(Continued)

   Theoretical categories, in contrast, place the coded data into a more general or abstract framework. These categories may be derived either from prior theory, or from an inductively developed theory (in which case the concepts and the theory are usually developed concurrently). They often represent the *researcher*'s concepts (what are called "etic" categories), rather than denoting participants' own concepts. For example, the categories "nativist," "remediationist," and "interactionist," used to classify teachers' beliefs about grade retention in terms of prior analytic distinctions (Smith & Shepard, 1988), are theoretical.

Most accounts of qualitative data analysis treat categorization as the fundamental activity in analysis, and the only one that involves manipulation of actual data, giving the impression that coding *is* qualitative data analysis. For example, LeCompte and Preissle stated that "the next step [after writing an initial summary] is to begin the time-consuming and laborious process of pulling apart field notes, matching, comparing, and contrasting, which constitutes the heart of analysis" (1993, p. 237). Numerous similar statements could be quoted (e.g., Hesse-Biber & Leavy, 2011, p. 309 ff.; Pfaffenberger, 1988, pp. 26–27; Tesch, 1990, p. 96; Webb & Glesne, 1992, pp. 796–801).

Tesch (1990, pp. 115–123) referred to this replacement of an original connected structure by a different, categorical structure as "decontextualizing and recontextualizing." She described recontextualizing as follows: "The [data] segment is settled in the context of its topic, in the neighborhood of all other segments of the data corpus that deal with the same topic" (p. 122). However, this new set of relationships is based on similarity rather than contiguity, and is thus not a "recontextualization" in the usual sense of "context," that is, a set of phenomena that are connected in time and space. These similarity-based relationships are quite different from a contiguity-based context, and confusing the two can lead to the neglect of actual contextual relationships. In addition, reordering the data in terms of particular categories can create analytic blinders, preventing the analyst from seeing alternative relationships in the data.

Both of these problems are illustrated by Atkinson's description (1992) of how his initial categorizing analysis of the teaching of general medicine affected his subsequent analysis of his surgery notes:

> [O]n rereading the surgery notes, I initially found it difficult to *escape* those categories I had initially established [for medicine]. Understandably, they furnished a powerful conceptual grid. Moreover, they exercised a more powerful physical constraint. The notes as I confronted them had been fragmented into the constituent themes. (pp. 458–459)

On returning to his original notebooks, Atkinson found that

> I am now much less inclined to fragment the notes into relatively small segments. Instead, I am just as interested in reading episodes and passages at greater length, with a correspondingly different attitude toward the act of reading and hence of analysis. Rather than constructing my account like a patchwork quilt, I feel more like working with the whole cloth. . . . To be more precise, what now concerns me is the nature of these products as *texts*. (p. 460)

Other researchers (e.g., Mishler, 1984, 1986) have also seen the neglect of context as a major defect of coding and other categorizing strategies. (This critique is similar to the realist critique of the "regularity" approach to causality, in terms of the latter's neglect of context as an essential component of causal explanation, discussed in Chapter 3.) Mishler argued that "the meanings of questions and answers are not fixed by nor adequately represented by the interview schedule or by code-category systems" (1986, p. 138), and claimed that systematic methods of narrative analysis are required to understand research interviews.

The categories thus generated are then usually linked into larger patterns (e.g., Tesch, 1993, p. 303 ff.); this subsequent step can be seen as contiguity-based, but the connections are made between the categories themselves, rather than between segments of actual data. However, using connecting techniques only on the categories, rather than the data, results in an *aggregate* account of contiguity relationships, and can never reconstitute the actual, diverse contextual connections that were lost during the original categorizing analysis (see the discussion of diversity in Chapter 4). Miles and Huberman warned that

> just adding up separate variables as in a quantitative survey approach will destroy the local web of causality, and result only in a sort of "smoothed-down" set of generalizations that may not apply anywhere in the real world of the sites. (1984, p. 151)

A common alternative to coding and thematic classification in qualitative research is the "case study" (Hesse-Biber & Leavy, 2011, pp. 255–276; Stake, 1995). In this approach, the unique context of each case is retained, and the data are interpreted within that context, to provide an account of a particular instance, setting, person, or event. However, case studies often employ primarily categorizing analysis strategies (e.g., Hesse-Biber & Leavy, 2011, p. 266; Merriam, 1988; Yin, 2003); their main advantage is that the categorizing (coding, thematic analysis, etc.) occurs within a particular case rather than between cases, so that the contextual relationships are harder to lose sight of. Qualitative case studies *can* be highly contextual or connected in their analysis (e.g., clinical case description), but are not inherently so.

Most qualitative researchers are aware of the dangers of decontextualization in using categorizing techniques. Works on qualitative methods are filled with warnings about context-stripping and admonitions to retain the connection of coded data with their original context. However, attention to context is often seen only as a *check* or *control* on the use of categorizing analytic strategies; most works say nothing about how one might *analyze* contextual relationships. For example, Lofland and Lofland argued that

> splitting the materials into mundane, analytic, and fieldwork files will facilitate staying "on top" of what is happening and evolving an analysis. But it also tends to obscure that nebulous quality called "context." . . . You should therefore keep a full set of your materials in the order in which you originally collected them. . . . [I]t is useful simply for reading and reviewing from beginning to end, as a stimulus to thinking about larger patterns and larger units of analysis. (1984, pp. 134–135)

Many other qualitative researchers also seem to view context as "nebulous" and resistant to systematic analysis. The development of explicit procedures for the analysis of contextual relations has, to a substantial extent, occurred in isolation from mainstream qualitative research, and it is to this approach that we now turn.

Connecting Strategies

Narratives, portraits, and case studies are often included in qualitative research reports as an accompaniment to categorizing analysis, and Barone (1990) argued that most qualitative texts are a mixture of narrative and paradigmatic (categorizing) design features. However, such uses of narrative are largely presentational rather than analytic; even Patton, who clearly used case studies as an analytic strategy (2001, p. 447), confounded this distinction by describing the case study as "a readable, descriptive picture of or story about a person, program, [or] organization" (2001, p. 450). Such presentational techniques partially compensate for the loss of contextual ties that results from a primarily categorizing approach, but they rarely are integrated with what is seen as the "real" analysis, or go beyond what is apparent in the raw data. Here, we are concerned with narrative or contextual approaches to data *analysis*, rather than simply presentation.

Connecting analytic strategies do not simply preserve data in their original form. Instead, they are ways to analyze and reduce data; this is generally done by identifying key relationships that tie the data together into a narrative or sequence, and eliminating information that is not germane to these relationships. Patton (2001, p. 447 ff.) and Seidman (1998) discuss the steps involved in selecting data to create case studies and "profiles," respectively.

The process of doing connecting analysis has received less attention than categorizing analysis. Narrative analysis is the most prevalent approach that has emphasized alternatives to categorizing analysis, but much of narrative research, broadly defined, involves categorizing as well as connecting analysis, and the distinction has not been clearly defined.

Many connecting strategies focus on the structure or significance of the narrative conveyed by the data (the latter usually being an interview transcript). These narrative strategies are informed by different disciplines. Discourse analysis, drawing from linguistics (e.g., Gee, 2005; Mishler, 1986), usually operates on a close, textual level where the semantic connections between different parts of the text are examined. One such strategy is the functional analysis carried out by Labov and his colleagues (1972, 1982; Labov & Fanshel, 1977; Labov & Waletzky, 1967). In their early work they focused on the temporal sequence of action within a narrative; in subsequent work, they attended to the larger, social meanings conveyed by the narrative structure.

A second kind of narrative strategy is informed more by sociology, anthropology, and clinical psychology. This approach is less concerned with the structure of the text and more with the meaning of that text for the participant. Seidman's "profiles" (1998) are one example of such an approach. In creating a profile from an interview transcript, Seidman first identifies and synthesizes the basic story line by reducing the text to those elements that are seen as important parts of the person's story. These segments are then crafted into a first-person account, normally (but not invariably) keeping the same order as these appeared in the transcript.

However, contiguity-based analytic strategies are not limited to narrative approaches. What Erickson (1992) called "ethnographic microanalysis of interaction" involves the detailed description of local interactional processes, and analysis of how these processes are organized (p. 204). The analytic process "begins by considering whole events, continues by analytically decomposing them into smaller fragments, and then concludes by *recomposing* them into wholes. . . . [This process] returns them to a level of sequentially connected social action" (1992, p. 217). Thus, instead of segmenting events and then *categorizing* these segments to create a structure of similarities and differences among these, this approach segments the data and then *connects* these segments into a relational order within an actual context.

Narrative strategies, as well as other connecting strategies, do not rely exclusively on contiguity. They utilize categorization, to a greater or lesser extent, to discern the narrative structure of the data (Linde, 1993, pp. 65–66). For example, identifying elements of plot, scene, conflict, or resolution in a narrative inherently involves classification. However, such classification is used to identify the elements of a narrative in terms of how they relate to other elements, rather than to create a similarity-based reordering of the data. Thus,

Mishler (1986) described some forms of narrative analysis that employ coding and categorization, but the categories he presented are *functional* rather than substantive categories. Such categories "provide a set of codes for classifying the 'narrative functions' of different parts of the account" (Mishler, 1986, p. 82), rather than constituting the basis for a reorganization of the data. Such categorization is a tool in contextual analysis, rather than a separate analytic process.

Narrative and contextual analyses, as strategies based primarily on contiguity rather than similarity, have disadvantages of their own. In particular, they can lead to an inability to make comparisons and to gain insights from the similarity or difference of two things in separate contexts. Some of our students have avoided coding their interviews or fieldnotes, because they felt that coding would destroy the contextual relations that they considered most important. However, when we persuaded them to try this, they said that it vastly increased their understanding of the data, because it broke them out of the fixed contextual frameworks within which they were working, and allowed them to see other relationships that they had been blind to. As one student wrote,

> At first, I resisted coding. The process seemed mechanical and reductive to me. I didn't want to violate the organic unity of my interviews, many of which had deeply moved me. To fracture these conversations into discrete pieces of information seemed like taking a pair of scissors and cutting up family photographs. However, as I started coding I soon realized what a powerful tool it was. To my amazement, I found connections between the interviews that I hadn't previously suspected. Not only did my informants share similar experiences, they sometimes used the same language to discuss those experiences. This was exciting, for I began to see that what had at first seemed like a mass of incoherent, intractable material did, indeed, have pattern and shape—and that in spite of the unique personalities and circumstances of my informants' lives, there were commonalities in both their experiences and the way they looked at things. (Huang, 1991)

An exclusive emphasis on connecting strategies can create what another student called an "imprisonment in the story" of a particular narrative—a failure to see alternative ways of framing and interpreting the text or situation in question. Wieviorka argued that comparison "may help deconstruct what common sense takes to be unique or unified" (1992, p. 170) and generate alternative perspectives.

Displays as Categorizing and Connecting Strategies

Displays (Miles & Huberman, 1994), as techniques for data analysis, can also be divided into similarity-based and contiguity-based forms. Miles and Huberman described a wide variety of displays, but most of these fell into two

basic types: matrices and networks (or figures). Matrices are essentially tables formed by the intersection of two or more lists of items; the cells in the table are filled with data, either raw or summarized, allowing comparison of the similarities and differences among the cells. The lists forming the matrix can be of individuals, roles, sites, topics, or properties of these, and can be organized in numerous ways, creating a large number of different types of matrices. Networks, on the other hand, are visual maps of the relationships (for Miles and Huberman, usually temporal or causal relationships) among individuals, events, states, or properties of these.

We see matrices and networks as, respectively, similarity-based and contiguity-based displays. Matrices are a logical extension of coding; they are created by constructing lists of mutually exclusive categories and then crossing these to create cells. Networks, on the other hand, are a logical extension of narrative or causal analysis, organizing events or concepts by time and by spatial or causal connection; they capture the contiguity-based relationships that are lost in creating matrices. Miles and Huberman provided examples of networks that link specific events, as well as those linking more abstract categories, although none were included that link actual data segments. They also presented a substantial number of hybrid forms that involve both categorizing and connecting strategies, such as time-ordered matrices and segmented causal networks.

Integrating Categorizing and Connecting Strategies

We have alluded to some of the advantages of combining categorizing and connecting strategies for analyzing qualitative data. However, even authors who explicitly discussed both types of strategies, such as Seidman (1998) and Atkinson (1992; Coffey & Atkinson, 1996), rarely address how to combine these. Implicitly, they seem to see the two as alternatives, or as parallel but separate analytic approaches.

While the separate use of the two approaches is legitimate and often productive, there are other possibilities as well. The most common is the sequential use of the two types of strategies, beginning with one and then moving to the other. For example, most qualitative researchers who employ coding strategies eventually develop a model of the connections or relational patterns among the categories, as mentioned earlier. However, this final step rarely involves direct analysis of data, and usually receives little explicit discussion. (Prominent exceptions are the work of Strauss, discussed shortly, and Miles and Huberman.)

Researchers who employ initial contextual or narrative strategies, on the other hand, often conclude by discussing similarities and differences among

the analyzed phenomena. (This is the reverse of the previous strategy of connecting categories into a relational sequence or network.) For example, Erickson describes the final step in ethnographic microanalysis as the "comparative analysis of instances across the research corpus" to determine how typical these analyzed units of interaction are (1992, p. 220). Again, however, this final categorizing step generally receives little explicit attention; an important exception is Miles and Huberman's discussion (1994, pp. 204–205; Huberman, 1989/1993) of a technique they called "composite sequence analysis," in which individuals' career trajectories are derived through connecting analysis of interviews, and these trajectories are then compared to identify common features and generate phases, subgroups, and modal sequences.

We suggest that it may be useful to think of this process in terms of categorizing and connecting "moves" in an analysis, rather than in terms of alternative or sequential overall strategies. At each point in the analysis, one can take either a categorizing step, looking for similarities and differences, or a connecting step, looking for actual (contiguity-based) connections between things.

A widely used approach to qualitative analysis that seems to us to employ this strategy is the "grounded theory" method (Strauss, 1987; Strauss & Corbin, 1990), although this aspect of Strauss's method of analysis has never, to our knowledge, been explicitly recognized. The initial step in analysis, which Strauss calls "open coding," involves segmenting the data, attaching conceptual labels to these segments, and making comparisons among the segments. However, the subsequent steps are predominantly connecting ones, despite being described as forms of coding; Strauss used "coding" to mean "the process of analyzing data" (Strauss & Corbin, 1990, p. 61). Thus, the next step, "axial coding," consists of

> specifying a category (*phenomenon*) in terms of the *conditions* that give rise to it; the *context* . . . in which it is embedded; the action/interactional *strategies* by which it is handled, managed, carried out; and the *consequences* of these strategies. (Strauss & Corbin, 1990, p. 97, italics in original)

This is almost a definition of what we mean by connecting analysis; the main difference is that these connections are to *categories*, rather than to data segments. The analytical steps subsequent to open coding involve making connections among categories, developing a "story line" about the central phenomena of the study, and identifying "conditional paths" that link actions with conditions and consequences. Confusingly, Strauss and Corbin referred to these connections as "subcategories," stating that "they too are categories, but because we relate them to a category in some form of relationship, we add the prefix 'sub'" (1990, p. 97).

Strauss continually integrated categorizing steps into these later stages, stating, for example, that "having identified the differences in context, the

researcher can begin systematically to group the categories. . . . This grouping again is done by asking questions and making comparisons" (Strauss & Corbin, 1990, p. 132). However, Strauss said very little about the grouping of *data* by category. Categorization, in the grounded theory approach, is manifested primarily in the development and comparison of concepts and categories. Nor does he deal with the analysis of specific contextual relations in the data, operating mostly in terms of relations among concepts.[2]

A similar alternation of strategies, but one that stayed much closer to the actual data, was employed by Miller (1991) in her study of adolescent friendships. We will therefore present an extended account of this study in order to illustrate one way in which these two strategies can be integrated.

We also have a second goal in presenting this example. The preceding account of the two different approaches to data analysis has been essentially categorizing rather than contextualizing; it classifies and compares the two types of strategies, rather than portraying them in any real context or discussing actual concurrent or sequential combinations of strategies. To complement this account, then, the following is a narrative or case presentation by Miller of her analysis. We hope that it will illustrate by example the difference between a contiguity-based account and a similarity-based one, and how to decide when to use categorical strategies and when to use connecting strategies. The account is written in the first person because it is a narrative of Miller's own struggle to make sense of her data.

A NARRATIVE EXAMPLE

"Once upon a time. . . ." While this account of data analysis is no fairy tale, it is nonetheless a story set in time and shaped by particular questions. Working with interviews with adolescents about their friendships, it seemed important to look closely at the features of the friendships, to understand in specific terms what they mean for the adolescents involved. In short, this seemed to call for a categorizing analysis, a close investigation of the components that seem to make up a relationship, for the purpose of investigating similarities across the friendships of different adolescents.

---

[1] Gerson (1991) addressed these issues in grounded theory research, using a distinction between heterogeneity heuristics and compositional heuristics that is similar to that between categorizing and connecting relations. Heterogeneity heuristics analyze similarities and differences among phenomena within a category, while compositional heuristics address the relationships among categories and between phenomena and categories. However, Gerson's compositional heuristics include relations that involve similarity as well as contiguity, such as taxonomic relations; the ones that most closely resemble what we call connecting relations are his part-whole and sequential relations.

I therefore began my analysis by formulating coding categories, coding the data, and constructing matrices. I coded the data for such elements as closeness, talk among friends, and dependence. These codes, for each interview, were then collected in matrices so that I could look across interviews for each concept. This helped me to focus on specific features of the data, informed by my research agenda as well as by the comments made by the adolescents themselves. With the completion of the matrices, though, two pressing issues emerged.

The first was that there was extensive overlap of data between the cells of the matrices. For example, many adolescents explained that part of being close to their friends involved talking with them. The matrix for closeness did not, however, capture the complexity of that talk, which involved information from other cells. These matrices seemed too simplistic for the complex, interconnected data I felt I had.

The second issue was that an essential aspect of the data was missing; namely, the narrative nature of the adolescents' accounts of their friendships. In their interviews, the adolescents did not offer isolated bits of information about their friends. Instead, what I heard were the stories of their relationships with their friends. As adolescents talked about their friends and explained why their friendships were important, they described their shared past and created a context from which to understand their relationship. This narrative quality of the data, and its implications for understanding their relationships, were lost in the process of coding and of creating the matrices.

To deal with the limitations of the matrices, and to capture the narrative quality of the data, I moved to what became the second phase of the analysis: the construction of narrative summaries. These summaries are narrative in that they seek to preserve the context and story of the relationship, yet they are summaries since they are my analytic abridgements of the narratives heard. These narrative summaries made use of extensive quotes from the data, but often involved a reorganization of the data to achieve what I, as the reader, perceived to be a concise account of the friendship narrative.

These narrative summaries were effective in holding on to the context as well as the story of the friendship. They did not, however, directly help me understand more clearly the meaning of that friendship experience for these adolescents. For that, I needed to look more closely at their relationships in light of my understanding of the larger context of that friendship. The next phase of my analysis, therefore, was to integrate the results of my categorizing and contextualizing strategies. This led to more depth within the concepts represented in the matrices; the category of closeness between friends, for example, was contextualized. By holding the narrative summaries against the matrices, I could track the meaning of closeness across different friendships for a particular adolescent or between adolescents, or trace its significance throughout a particular friendship.

Data analysis had become, for me, an iterative process of moving from categorizing to contextualizing strategies and back again. My understanding of the narrative context of the friendships informed my interpretation of the particular concepts and categories I had identified as important in these adolescents' friendships. At the same time, the particular concepts I focused on in the categorizing analysis allowed me to look at the narratives in new ways, and to see contextual relationships that were more complex than the temporal ordering of events within the narratives. My understanding of the meaning and experience of friendship for these adolescents was not stripped of the context, which the adolescents provided, nor was it locked into and limited to individual friendship stories. Coding and matrices were combined with narrative summaries to achieve an understanding of the interviews that neither could have provided alone.

A final example of a strategy that we see as combining categorizing and connecting "moves" in analysis is the "listening guide" strategy (Gilligan, Spencer, Weinberg, & Bertsch, 2003) for analyzing interviews. This strategy, which the authors describe as a "voice-centered relational method," involves a sequence of readings (the authors use the term "listenings") of the interview transcript, each focused on a different aspect of the speaker's expression of her experience within the context of the research relationship. This approach is premised on the idea that a person's voice is polyphonic rather than monotone, that different "voices" can be identified within an interview.

The first listening is typically for the "plot" of the interview—what stories are being told, and in what contexts—and the researcher's response to these. The second listening is for the voice of the "I" who is speaking—how does this person speak about himself or herself? This involves underlining all passages containing an "I," along with the associated verb and any other important words, and creating a separate text with only these segments, keeping them in their original order. Subsequent listenings depend on the specific purposes and questions of the research, but are typically "contrapuntal," focusing on contrasting issues and "voices." In Gilligan's original use of this method, the focus was on differences between men's and women's moral judgment, and the listenings were for the voices of justice and of care, and of a separate and a connected self. These later listenings are not necessarily specified in advance; they may be inductively developed. The final step is to pull together an interpretation of this person's perspective on these issues.

The listening for different "voices" is clearly a categorizing move in that it identifies segments that are similar in some way—that they are first-person statements, or deal with a particular issue. However, in contrast to traditional coding, these segments are not fragmented and reorganized by topic; they are kept in sequential order. In the case of "I" statements, these segments are used

to create an "I poem" that captures the "associative stream of consciousness" (Gilligan et al., 2003) running through the interview; this is a connecting step in analysis. As in Miller's example, the analysis is composed of a mix of categorizing and connecting moves, with each strategy compensating for the deficiencies of the other. For example, the "I poem's" initial categorizing step allows the listener to focus specifically on the voice of "self" without the interference of extraneous material, foregrounding this aspect of the narrative, while the connecting step of preserving the chronological order of the statements allows the listener to follow the sequential links between these statements.

Like Atkinson (1992; Coffey & Atkinson, 1996), we see categorizing and connecting approaches as inherently complementary strategies for data analysis. The complementarity of similarity and contiguity relations in language is generally recognized, and is a central theme in the paper by Jakobson (1956) cited earlier. Jakobson, who played a major role in developing the similarity/contiguity distinction with reference to language, also applied this distinction to aphasia, the loss of ability to understand or express speech, caused by brain damage. He identified two types of aphasia, which he labeled *similarity disorder* and *contiguity disorder*, each reflecting a loss of one of these two essential components of language. A key point in Jakobson's argument is the *complementarity* of the two dimensions; he argued that the loss of either dimension resulted in an inability to use language effectively.

It seems to us that the defects in qualitative analysis that result from ignoring one or the other of these relationships among qualitative data are analogous to the communicative disturbances Jakobson (1956) described in these two types of aphasic patients. The use of computers in qualitative research is a case in point.

Computers and Qualitative Data Analysis

Computer programs for analyzing qualitative data have had a major influence on how analysis is done (Hesse-Biber & Leavy, 2011; Richards, 2005; Weitzman & Miles, 1995), and will undoubtedly have even greater impact in the future. However, so far computers have been used primarily for categorizing rather than connecting types of initial data analysis, due to the ease and power with which computers can perform similarity-based functions such as sorting and comparison.

There is thus a danger that, following what Kaplan (1964, p. 28) called "The Law of the Instrument," the ease of using computers for categorizing analysis will reinforce this approach and lead to the neglect of connecting strategies. Pfaffenberger argued that

> a technology is like a colonial power—it tells you that it is working in your
> best interests and, all the while it is functioning insidiously to dim your criti-
> cal perception of the world around you. You will remain its victim so long as
> you fail to conceptualize adequately what it's up to. (1988, p. 20)

Against this view, Tesch claimed that "computers, like scissors, are tools.
In themselves, they have no influence on the research process" (1993, p. 279).
This claim neglects the ways in which the *decision to use* particular tools, such
as the decision to use scissors rather than some other technique, involves
assumptions, often unconscious ones, about the nature of analysis (for
example, that it begins with "cutting up" the data). Such tools privilege certain
analytic strategies and inhibit others.

For example, Agar (1991, p. 181) was once asked by a foundation to
review a report on an interview study that it had commissioned, investigating
how historians worked. The researchers had used the computer program "The
Ethnograph" to segment and code the interviews by topic and collect together
all the segments on the same topic; the report discussed each of these topics,
and provided examples of how the historians talked about these. However, the
foundation felt that the report hadn't really answered its questions, which had
to do with how individual historians thought about their work—their theories
about how the different topics were connected and the relationships they saw
between their thinking, actions, and results.

Answering the latter question would have required an analysis that elu-
cidated these connections in each historian's interview. However, the catego-
rizing analysis on which the report was based fragmented these connections,
destroying the contextual unity of each historian's views and allowing only a
collective presentation of shared concerns. Agar argued that the fault was not
with The Ethnograph, which is extremely useful for answering questions that
require categorization, but with its misapplication. As he commented,

> I don't mean to pick on The Ethnograph. On the contrary, later I describe
> a study where, if The Ethnograph had been available, I would have been the
> first in line. I do mean to say that a program like The Ethnograph represents
> a *part of* an ethnographic research process. When the part is taken for the
> whole, you get a pathological metonym that can lead you straight to the right
> answer to the wrong question. (1991, p. 181)

Connecting uses of computer software do exist. So-called "theory-
building" programs (Weitzman & Miles, 1995) can use connections between
categories to assist in testing hypotheses about relationships and establishing
typical sequences. However, these uses are almost always based on a previous
categorizing analysis, and the connecting functions focus on linkages between
concepts, as in Strauss's "grounded theory" approach to analysis, rather than

on linking actual data. An exception is Richards's (2005) description of software for establishing links among data and data files. However, she focused mainly on links between different *types* of data, such as between fieldnotes and memos, and on links between *different* interviews or observations, as well as links between data categories. Her emphasis was almost entirely on categorizing analysis, and she did not discuss linking data *within* a specific context, or identifying relationships of contiguity rather than similarity/difference.

Despite this, there are ways that computers can be used to assist in the direct connecting analysis of qualitative data. One way is to mark, extract, and compile selected data from a longer text, simplifying the task of data reduction in producing case studies, profiles, and narratives. This is a function that any word processor can perform, but one that could be improved by software specifically designed for this purpose. Another is to use graphics programs (such as Inspiration) to develop network displays of events and processes (Miles & Huberman, 1994). So-called "hypertext" programs (Coffey & Atkinson, 1996, pp. 181–186; Dey, 1993, pp. 180–191) allow the user to create electronic links among any segments, within or between contexts; a few of the more structured "theory-building" programs, such as ATLAS/ti and HyperRESEARCH, will not only do this, but will display the resulting networks. Software that is designed to facilitate such strategies could move case-oriented, connecting analysis beyond what Miles and Huberman (1994) call "handicraft production," and could help to prevent the "pathological metonym" that Agar warns against.

In summary, we have argued that the distinction between similarity-based (categorizing) and contiguity-based (connecting) analytic strategies, which we see as grounded in a realist understanding of relationship types, is a useful theoretical tool, both for understanding how qualitative researchers analyze data and for improving our analyses. The two strategies, rather than being antagonistic and mutually exclusive alternatives, are best seen as complementary and mutually supporting, each having its own strengths and limitations. We think that Wieviorka's statement about sociological and historical analyses (which we see as analogous to categorizing and connecting strategies) also applies to qualitative data analysis: that "research will advance not by confusing but by combining these approaches" (1992, p. 163).