A document you requested through Joyner Library’s Interlibrary Loan or Document Delivery service follows. For Interlibrary Loan requests, the ILLiad system received the document from an outside provider and automatically emailed it to you without staff review.

If pages are missing or are illegible, if wrong pages are attached, or if other problems exist with the document, please contact the Interlibrary Loan staff immediately for a replacement.

Please disregard any library work forms that might also be attached to this document. Some forms might indicate that payment is due for the document; be assured that any fees due are being paid by Joyner Library.

The notice below serves to remind you that you must comply with Copyright Law in your use of the following document. Visit Joyner Library’s Copyright Research Guide to learn more about copyright; questions may be directed to ECU’s Copyright Officer, Joseph Thomas.

If you have compliments, suggestions, or complaints related to Joyner Library’s Interlibrary Loan and Document Delivery services, please share them with us.

Thank you for using Joyner Library.

Rebecca Harrison, Borrowing Manager
Andrew Morgan, Lending & Document Delivery Manager
William Gee, Access Services Librarian

US COPYRIGHT NOTICE:

NO further reproduction or distribution of this copy is permitted by electronic transmission or any other means. The user should review the copyright notice on the following scanned image(s) contained in the original work from which this electronic copy was made.

SECTION 108: United States Copyright Law

The copyright law of the United States [Title 17, United States Code] governs the making of photocopies or other reproductions of copyrighted materials.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “Fair Use,” that use may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law. No further reproduction and distribution of this copy is permitted by transmission or any other means.
| **Rapid #:** | -9585893 |
| **CROSS REF ID:** | 599681 |
| **LENDER:** | IYU :: Main Library |
| **BORROWER:** | ERE :: Joyner Library |
| **TYPE:** | Book Chapter |
| **BOOK TITLE:** | Digital writing research: technologies, methodologies, and ethical issues / edited by Heidi A. McKee, Danielle Nicole DeVoss. |
| **USER BOOK TITLE:** | Digital Writing Research: Technologies, Methodologies, and Ethical Issues |
| **CHAPTER TITLE:** | Coding Digital Texts and Multimedia |
| **BOOK AUTHOR:** | Blythe |
| **EDITION:** | |
| **VOLUME:** | |
| **PUBLISHER:** | |
| **YEAR:** | |
| **PAGES:** | 203-227 |
| **ISBN:** | 9781572737068 |
| **LCCN:** | |
| **OCLC #:** | |

Processed by RapidX: 8/27/2015 4:32:22 PM

This material may be protected by copyright law (Title 17 U.S. Code)
CALL #: PE1478 .D54 2007
LOCATION: IUY :: Main Library :: Moody General Collection
TYPE: Book Chapter
BOOK TITLE: Digital writing research: technologies, methodologies, and ethical issues / edited by Heidi A. McKee, Danielle Nicole DeVoss.
USER BOOK TITLE: Digital Writing Research: Technologies, Methodologies, and Ethical Issues
CHAPTER TITLE: Coding Digital Texts and Multimedia
BOOK AUTHOR: Blythe
EDITION: 1st
VOLUME: 203-227
PUBLISHER:
YEAR:
PAGES:
LCCN:
OCLC #: IUY OCLC #: 85765972
CROSS REFERENCE ID: [TN:599681][ODYSSEY:150.216.68.52/ERE]
VERIFIED:

BORROWER: ERE :: Joyner Library

This material may be protected by copyright law (Title 17 U.S. Code)
Codings, classifying, and coding data allow researchers to identify a set of artifacts, to define a unit worth analyzing within them, to create codes for classifying instances of that unit, and to extrapolate based on the results of such coding. Digital artifacts pose interesting coding-related issues because they are less stable than print artifacts, alter relations between creator and audience, and can incorporate multiple media. This chapter addresses the challenges of coding digital data by discussing the procedures and assumptions that inform many data-coding activities and by exploring ways that such activities may be adapted for digital writing research. This chapter addresses both procedures and assumptions and draws examples from a study analyzing how colleges and universities use the World Wide Web to promote campus-community partnerships.
INTRODUCTION

Many of the questions we ask about digital writing require us to look for patterns in large amounts of data. If we wanted to explore social relations on an email discussion list, to know how a group of people represents themselves on the Web, or to compare uses of written and visual communication in multimedia instructional texts, then we would need a way to sort and classify the phenomena we encountered. Let's say you were curious about how racial and ethnic diversity was being represented in the many Web sites and CD-ROMs that now supplement college-level composition texts. You could devise a way to count the number of times males and females of various races and ethnicities were depicted in these resources. Such depictions could come from gender-, ethnicity-, and race-based references in the written portions of the sites; audio spoken by a male or female voice, with a dialect that marks the speaker as a member of some group; and photos, drawings, and video that depict males or females of various races and ethnicities. You could also devise a way to classify the activities being depicted in these media. How often are different people shown writing, recreating, collaborating, or volunteering? How often are different people depicted as leaders, as followers, as effective writers, as ineffective writers, and so on? What might all this suggest regarding publishers and prevailing views of diversity? Such a study would involve numerous variables and large amounts of data. How would you handle such a quantity and range of phenomena? For that matter, how would you determine what was worth classifying?

The act of sorting and classifying can be called data coding. Coding requires researchers to identify a set of artifacts (e.g., written documents, photographs, videotaped interactions), to define a unit worth analyzing within them, to create codes for classifying instances of that unit, in many cases to test the reliability of that work, and to make these decisions and actions public. The process of coding ideally provides a way to measure the prevalence of some set of phenomena. Thus "coding schemes," as Roger Bakeman (2005) said, "can be thought of as measuring instruments just like rulers and thermometers" (p. 375). Although many resources exist on sorting and classifying verbal data,¹ digital artifacts present new challenges because they are less stable than printed artifacts, alter relations between creator and audience, and can incorporate multiple media. This chapter addresses these new challenges by discussing the procedures and assumptions that inform many data-coding activities and by exploring ways that
such activities may be adapted to digital writing. Because this chapter addresses both procedures and assumptions, it is divided into two major sections: method and methodology. I have adopted this organizational pattern because how you approach any coding activity depends in part on what you want to explore and your assumptions regarding research and the epistemological status of verbal and visual language. If you already understand methods of data coding, you may want to skim the sections "Methods of Coding Data" and focus on "From Method to Methodology." If you are unfamiliar with data-coding methods, or if you seek tips for coding the variety of media presented in digital writing, begin with "Methods of Coding Data."

Writers such as Sandra Harding (1987), Gesa Kirsch and Patricia Sullivan (1992), and Sullivan and James Porter (1997) distinguished between method (procedures, techniques) and methodology (assumptions, theories). Sullivan and Porter (1997) wrote that method can be equated "with particular observational procedures or data-collection strategies and with specific data analysis techniques" (p. 11). The first portion of this chapter is focused on method because it offers an introduction to techniques of data coding and advice on adapting that method to digital artifacts. The set of strategies and techniques discussed in this chapter are used in various social scientific methods such as content, discourse, and genre analysis. Moreover, they are used in a range of disciplines, including linguistics, communication, writing studies, and psychology. These methods have been used to study a wide range of phenomena, including interview transcripts, chat sessions, Web sites, banner ads, television shows, and videotaped interactions.

The second portion of the chapter is focused on methodology. As Kirsch and Sullivan (1992) stated, methodology is "the underlying theory and analysis of how research does or should proceed" (p. 2). Methodology is worth discussing because most data-coding techniques were developed within a modernist context of truth and objectivity, but have been adapted by researchers who have rejected modernist claims. Sullivan and Porter (1997) differentiated, for instance, between a modernist approach to Theory ("the traditional kind that claims universal authorizing status") and a more postmodern approach to theorizing ("local, contingent, and situated reflection and analysis that has a status as heuristic rather than as Law"; p. 102). Some who still accept modernist assumptions regarding truth and objectivity have focused on ensuring scientific rigor through objectivity, validity, and reliability when coding digital writing (see, for example, McMillan, 2000; Potter & Levinne-Donnerstein, 1999; Weare & Lin, 2000). Others have argued specifically for the value of data coding given postmod-
ern assumptions (see, for example, Barton, 2002). In this chapter, I argue for the continuing value of data coding for theory rather than Theory; I explore this theme by discussing the epistemological status of texts and the value of data coding in critical research.

**METHODS OF CODING DATA**

Data coding appears in several variations, including content, discourse, genre, and rhetorical analysis. Because these variations share a core set of steps, this chapter is focused on common features rather than on differences. This overview of core steps should provide a manageable introduction to data coding and the ways it may be adapted to digital environments. Because such environments increasingly present a range of media, this chapter draws ideas and suggestions not only from writing studies, but also from fields such as psychology, communication, and journalism—all of which have a history of coding nonverbal phenomena.

Two caveats before we proceed: First, although the print medium requires me to present the following steps in one order, they are not necessarily sequential. A researcher may define a unit of analysis, for instance, before selecting a sample. Also, units of analysis and coding schemes may be revised as a project progresses. As with writing, data coding can be a recursive process. Second, the purpose of the following sections is to provide an overview of possibilities, but not to offer canned methods. As Cindy White and Jack Sargent (2005) wrote, "ready-made research solutions are rare" (p. 3); researchers must develop codes and units to fit their circumstances. I hope to aid such development by offering ideas, examples, and references to other sources. Moreover, because multimedia can present writing researchers with the challenge of coding nonverbal phenomena, I draw from various traditions (such as psychology and communication) that have well-established coding practices for such types of texts.

**Defining Sets and Selecting Samples**

Any research into communication requires an appropriate group of texts for study. Texts are necessary even in studies of spoken communication, where researchers must somehow record what was said before they can analyze it. To identify such a group, you should: (a) define a set of poten-
tial texts, and (b) select a sample from that set. The type of set you define depends on the kind of question you wish to ask. Paul Kei Matsuda (2002), for example, noted that most studies of identity and power in online discourse were focused on speakers of English. Because he believed that “conclusions drawn from the perspective of only one language may not be applicable to other languages,” he analyzed “about 200 e-mail messages that were exchanged among the members of TESOL Link between July and September 1997” (pp. 40–41). Recently, to provide another example and one I focus on later in this chapter, I wanted to know how colleges and universities were using the World Wide Web to promote campus–community partnerships; such partnerships are often set up by colleges and universities as a way to foster civic participation, urban problem solving, and service learning (Blythe, 2004). I wanted, in part, to know whether institutions were using the Web merely to promote their services, or whether and how they were using networked computers to do community work (e.g., by constructing community databases or sponsoring online forums). The data set in this case included all campus-sponsored Web sites in the United States devoted to campus–community partnerships.

Once you have identified a set of potential texts, you need to select a sample from it. In some projects, the set you define will be small enough to allow you to code every item, which would be called comprehensive sampling. In other cases, however, the set you define may be too big to code in its entirety. You may have to select a sample from that set. Samples may be selected in several ways; for example:

- With convenience sampling, you select a set of texts because they are readily available. Although this sampling method is easier than others, it carries little credibility on its own and should not be the sole method of selection. With increased levels of access provided by the Internet, many texts are conveniently available. That was certainly the case when I studied ways that campuses were using the Internet for community outreach. Although convenience can play a role, samples should be selected through a mix of methods.

- With criterion sampling, you select a set of texts because they meet certain criteria, which may be defined according to such factors as textual features, author attributes, intended audience, or types of media. For example, you may want to sample texts that present a best, typical, or worst case of the phenomenon you wish to study. (How you decide what is best, typical, or worst may depend upon how others have defined the terms. In other
words, background reading is crucial to setting criteria.) Or you may want to sample texts written by students at certain types of universities, or by people who have been maintaining blogs for more than 2 years, or by female students to female instructors. Or you may want to select projects that incorporate at least three types of media (e.g., audio, text, and animation; or text, audio, and video).

- With random sampling, you assign a number to each element in your overall set and then use a random number generator or chart to guide the selection of elements from that set. For example, if you were studying the transcripts of student chat sessions during collaborative projects, you could assign a number to each session and then use a random number generator to select a set of sessions to code.

Because of the scope and mutability of the Web, you may need to use a variety of sampling methods. For example, in my study of campus-community partnerships, I followed the advice of Christopher Weare and Wan-Ying Lin (2000) for seeking publicly available sites on the Internet. I consulted a “collector site” (i.e., an index of campus outreach grant winners maintained by the Department of Housing and Urban Development), a search engine (Google.com), and two people involved in such partnerships. I consulted these different resources because any one method would have presented a skewed list—Google.com presents sites based on popularity, and the collector site presented only grant winners. From these resources, I was able to build a set of sites and then use a mix of criterion and convenience sampling to select my sample.

Even after a sample has been selected, digital writing presents additional challenges. One primary challenge involves defining boundaries. In print texts, boundaries are usually easier to find. We can tell where a journal article begins and ends. Boundaries are less obvious, however, when studying hypertexts such as Web sites or blogs. Which links should be followed and which ignored? One way to alleviate this conundrum is to rely on the domain name system. You may decide to look only at pages from, say, www.school.edu/class1/. You'd examine www.school.edu/class1/page1.htm and www.school.edu/class1/page3.htm but not www.school.edu/class2/page1.htm (because the latter comes from “class 2” rather than “class 1”)—unless, of course, you decide to code all pages on www.school.edu (for additional advice on sampling in networked environments, see Weare & Lin, 2000). Although this technical approach to defining a sample may be helpful, it cannot be relied upon exclusively. For example, I noticed
that one campus–community outreach project I was studying presented reports and promotional information under one domain name and a community database under another. In that case, I had to decide whether to include both sites.

Another challenge for the analysis of digital writing involves the instability of many digital texts. Parts of a sample that you decide to code on one day may have been revised by the time you actually do your coding. You may want to gather your own copies of the texts you intend to code, making sure to note the dates on which you made your copies. You may even want to insert your own information into downloaded and stored Web pages by using the <i>....</i> tag in HTML. For example:

<i>....downloaded from www.school.edu/class1/page1.htm on 2004.10.15....</i>

This tag lets you keep notes underneath the page without affecting how the page is presented in a Web browser.

**Defining a Unit of Analysis**

A unit of analysis is the clearly defined phenomenon that you wish to study. Geisler (2004) called it “the level at which the phenomena of interest occurs” (p. 29). Units can range from small (such as words) to large (such as entire texts); they can be verbal, visual, aural, and even what I would call “technical,” such as links in a hypertext. This section describes points along that range.

**Verbal Units in Written and Spoken Language**

Researchers have examined words at all levels—from small linguistic units to longer rhetorical statements. The list below, though not comprehensive, provides a sense of the range of verbal units that may be coded:

- **Words, Phrases, or Clauses.** For example, you may want to code indexical words (*I, she, here, there, now, then*) to see how a writer orients her reader to other phenomena. Or you may want to code nouns and nominal phrases as a way to identify the metaphors people use when describing computer technology (e.g., *master, slave*). Or you may want to code clauses that contain modals to determine the extent to which a group communicates probability (*We might do that*), advisability (*You ought to do that*), or condi-
tionality (*We would have done that*). Matsuda (2002), for example, looked for "social relation markers—formal verbal endings, address terms, and honorifics" (p. 46) as evidence of power relations between members of an email list with the goal of providing "a Japanese-language medium list for English teachers in Japan who pursued graduate studies in TESOL or related fields in the United States" (p. 41).

- **T-Units.** As Geisler (2004) wrote, a t-unit "consists of a principle clause and any subordinate clauses or nonclausal structures attached to or embedded in it" (p. 31). T-units are worth coding when you want to examine the kinds of moves people make in language because such units contain relationships between people or things (nominals) and actions (verbals). You may want, for instance, to count the number of times people pose questions, make assertions, or give directions. Yi Yuan (2003), for example, examined the number of times two non-native speakers of English "repaired" their statements in chat room sessions, either by "seek[ing] remedies or confirmation" from Yuan or by "offer[ing] solutions or suggestions themselves to rectify the errors" (p. 198).

- **Exchanges.** Exchanges involve the give-and-take between two or more people. For example, an exchange occurs when a writing center consultant poses a question and a student responds. If you were coding exchanges, you would apply one code to the entire exchange. In the case of email correspondence, you might apply one code to two messages, the query and the reply. Although the boundaries of exchanges are relatively easy to spot in many circumstances, they would be more elusive in a transcript from something like a MUD/MOO session with four or five participants. Because person A may respond to person C just as person B is responding to person D, a chat transcript often presents a jumble of exchanges. Coders have to determine where the exchanges occur.

- **Rhetorical Units.** A rhetorical unit is a segment that can be classified as one type of rhetorical move—a move with the same author, intended audience, and purpose. More than any other type of analytical unit, a rhetorical unit can vary in length. An entire Web page may be devoted to sharing contact information. Or an entire email message may be devoted to soliciting one type of response from a writing center consultant. Or, conversely, one
type of Web page or email message may contain several moves. Susan Herring (1996), for example, studied a type of rhetorical unit called the macrosegment, which had such features as a coherent purpose, a consistent set of linguistic choices, and sometimes visual cues such as blank lines or indentation. By studying these units in two email list exchanges, Herring questioned the traditional assumption that women communicate for interpersonal reasons but men communicate to exchange information.

**Nonverbal Units in Spoken Language**

Whereas verbal units help researchers explore how meaning is conveyed through what is said or written, nonverbal units help us explore how something is communicated. Multimedia projects that incorporate significant portions of audio or video may merit the exploration of nonverbal features. If your interest focuses on multimedia projects with audio, you may want to learn how others have coded spoken language. Speakers can convey meaning using a range of gestures and nonverbal features such as loudness (amplitude), pitch (frequency), and rate of speech (often measured as syllables per minute). Consider, for example, how changes in pitch can add variety to the sentence *Were you there last night?* The meaning would change if the speaker changed pitch for various words in the sentence. Emphasis on the word *you* (*Were you there last night?*) would suggest that the speaker knew when an event had happened but wondered whether another person had attended; emphasis on the word *night* (*Were you there last night?*) would suggest that the speaker was uncertain when an event happened. Such nonverbal features of speech have been shown to have social significance because they reinforce expressions of emotion, attachment, and social influence (see Buller & Jones, 2005; Carrère & Gottman, 2005; for a brief history of the measurement of nonverbal features of speech as well as information on software that can be used to aid such coding, see Tusing, 2005).

Physical phenomena such as gestures and facial expressions constitute another form of nonverbal communication. If your interest focuses on multimedia that contains significant amounts of video depicting people, then you might draw from psychology to code behaviors depicted. For example, the Specific Affect Coding System (SPAFF) "is a gestalt system of observation that integrates non-verbal and physical cues, voice tone, and speech content to identify specific affects" (Jones et al., 2005, p. 164). This system has been used to code interpersonal behaviors such as touching,
physical distance, body orientation, gazing, nodding, smiling, and furrowing of the eyebrows (for an extended list of codes for such behaviors, see Guerrero, 2005, pp. 225–230). Another approach to examining such phenomena is tie-sign typology. As Walid Afifi and Michelle Johnson (2005) explained, tie-signs are “actions that provide evidence of a personal relationship” (p. 189). Such evidence includes handshakes, arm links, kissing, and various types of embraces.

Coding Visuals

In the previous two sections, I distinguished between verbal (what) and nonverbal (how) units of written and spoken language. A similar distinction can be made regarding the coding of visuals such as photographs, drawings, and animation. In this section, you will find advice on coding the content of an image, on what appears there; in the next, you’ll find advice on coding the way an image is constructed, on how it was designed to present content. When it comes to coding what is portrayed in a visual, researchers have created units to help them isolate and count the appearance of certain types of images. For example, Ken Smith and Cindy Price (2005) developed a coding scheme that let them study the ways members of different races were depicted in photojournalism. They created codes to identify the race of the people in the photos (black, white, and other) and the subject matter (e.g., politics, education, sports, religion). Then they examined photographs in 500 “general circulation nondaily newspapers in the United States” (p. 131). Their coding scheme let them explore whether blacks are represented fairly in nondaily newspapers.5

As Smith and Price’s (2005) study illustrates, codes can be developed to identify visual phenomena that carry social meaning—phenomena such as skin color, clothing, artifacts, setting, and so forth. Researchers who wish to develop a set of culturally significant codes might want to explore the work of Sandra Moriarty and Lisa Rohe (2005) who wrote about “cultural palettes” in advertising. A cultural palette is “a set of culturally sensitive symbols and colors, as well as other graphic elements such as layouts and artistic styles, that may reflect cultural nuances” (p. 119). For example, an appropriate cultural palette for Korean students may include symbols (tiger, dragon, bamboo, but not cat, cowbird, snake) and colors (white, blue, green, but not purple, pink, black). Although the cultural palette concept was developed to aid advertisers, it could be adapted for coding. That is, one could code for types of colors, types of symbols, and types of activities that appear in a set of multimedia projects.
Coding and Aesthetics

In addition to coding what is depicted in multimedia, there are ways to code how visual elements are constructed. Aesthetics theory may provide a set of concepts to aid such an examination. As Herbert Zettl (2005) wrote, “media aesthetics examines five basic aesthetic image elements that provide the aesthetic materia—the raw material—of television, film, and computer-generated images” (p. 366). Those elements are:

- **Light and color.** Consider whether light is directionless (coming from no particular source and casting no shadows) or highly directional (coming from an obvious source and casting shadows). If the light is highly directional, you could code the location of the source. Does the light come from above, below, to the left? Or you might code the amount of light (brightness versus darkness). When it comes to color, remember that colors are associated with certain feelings and meanings. (See Moriarty & Rohe, 2005, whose work was discussed in the previous section.)

- **Two-dimensional space.** Examine the way multimedia designers use aspect ratio (the relation between picture width and height) and screen size. Are designers trying to reproduce the feel of a movie, book, television show? You might develop a code for the way designers frame certain elements within that space. Zettl (2005) wrote about vectors, which he described as “forces with a direction and magnitude (relative strength)” (p. 372). Graphic vectors are created by a collection of elements in an image; index vectors “are created by somebody or something pointing unquestionably in a specific direction” (p. 372). What kinds of dominant lines are created in an image? Is a person looking directly at the camera or to the right or left?

- **Three-dimensional space.** Examine whether images are presented up close (as with a zoom lens) or further away (as with a wide-angle lens). You might examine whether a drawing or photograph is intended to present detail or a sense of intimacy and intensity, or whether the visual is intended to present a panorama or a group shot that loses individual detail. For example, Robert Tiemens (2005) coded the number of times that four major networks used wide-angle, medium, close-up, and extreme close-up shots of Jesse Jackson’s speech at the 1988 Democratic national convention. Tiemens wanted to code these
shots because the way an image is framed "directs the viewer's attention to the speaker or to pertinent details of the event" (p. 394). A close-up draws attention to the speaker while a wide-angle shot commands less attention.

- **Time-motion.** If you were coding multimedia that included animation or video, code for changes over time. For example, is the video live or edited after-the-fact? If the video was edited after-the-fact, or if you were looking at animation, then you might code how one sequence transitions into another. Such codes could be drawn from film and television studies. You might code for cuts, dissolves, zooming in or out, or panning (see Tiemens, 2005, p. 190). For other types of codes relating to time-motion, see the section below on coding transitions.

- **Sound.** Code types of sounds (e.g., sounds from nature versus sounds from urban spaces, upbeat music versus sad music). You might also code the relationship between what one hears and what one sees or reads. For example, does the background music in a multimedia piece reinforce or juxtapose with the words in the piece?

Just as nonverbal cues can affect the reception of verbal data, so can media elements affect perception. Such elements, argued Zettl (2005), "are primarily responsible for providing the all-important background against which we tend to interpret all the literal aspects of the event" (p. 366).

**Coding Links**

If you are studying hypertext, you may want to examine ways that transitions and cohesion are created through links. For example, you might use the "hyperphoric grammar" developed by David Norton, Beverly Zimmerman, and Neil Lindeman (1999). Relying on the work of M. A. K. Halliday and Ruqaiya Hasan (1976), Norton et al. developed a classification scheme that described whether links were sequential (chronophora), referential in relation to a starting point or location (primaphora, such as a link back to a home page), nonsequential (paraphora), or related to a computer action (compuphora, such as an email link; they offer a helpful table that summarizes and illustrates types of links; see p. 187). If you were studying multimedia and narrative, you might adapt Scott McCloud's (1994) typology of transitions between panels in comic strips. McCloud identifies six types of transitions: With moment-to-moment transitions, very little time
has elapsed between each panel. A blink may be represented in two panels. In the first a girl may have her eyes open, and in the second she may have them shut. With action-to-action transitions, more time has elapsed between each panel. One panel may depict a pitcher pitching a ball, and another may depict a batter hitting it. With subject-to-subject transitions a sequence of panels “stay[s] within a scene or idea” (p. 71). One panel may show a runner crossing a finish line, and another may show a hand stopping a stopwatch. Scene-to-scene transitions represent significant shifts in time and space. This may include a shift from one city to another or from action outside a building to action inside. Aspect-to-aspect transitions ignore time to show “different aspects of a place, idea or mood” (p. 72). A sequence of panels may, for instance, show different views of the same room in the same instance of time. Finally, non-sequitur transitions present a sequence of panels with no clear relationship of time, place, or idea.

Coding for Manifest or Latent Content

The range of units described in the previous set of sections can be split into two subsets: manifest units and latent units. Manifest units are observable phenomena in a text. Such codes are typically lexical or syntactic, such as words, phrases, clauses, and t-units (Grant-Davie, 1992). Consequently, manifest content is relatively easy to spot and can be measured quantitatively. You could even use the search function in word-processing software or a Web browser to find words and phrases. Latent content, on the other hand, “shifts the focus to the meaning underlying the elements on the surface of a message” (Potter & Levine-Donnerstein, 1999, p. 259). Researchers who examine latent content look either for the purposes to which language is put or for the schema that such language evokes in readers (a measure of reader response). Rhetorical units are examples of units of latent content because they require the coder to make an interpretation, to infer purpose, and to decide when purpose or audience has shifted. To code for latent content requires a “sensitivity to context,” and an ability to understand the purpose, audience, and other factors that motivate any statement (Huckin, 2004, p. 15). For example, I examined latent content in the studies of campus-community partnerships. I had to infer the purpose that faculty and staff had for creating content for their Web sites. Were writers advertising their services, providing raw data or analysis, or promoting online collaboration? Such judgments cannot always be made easily, as is illustrated in the next section.
Creating a Set of Codes

As was mentioned previously, data coding is more recursive than linear. Consequently, when and how a set of codes is defined varies from study to study. Some researchers come to a project with a set of codes in mind. They may have defined a set for themselves, or they may have adopted an existing set. Milena Collot and Nancy Belmore (1996) adopted an existing set, for example, when they studied “electronic language.” For them, the unit they studied was the word, which they coded grammatically—nouns, prepositions, pronouns, adjectives (that is, they searched for manifest content). Using the codes from Douglas Biber’s (1988) multidimensional-multifeature model, Collot and Belmore were able not only to identify significant features of language used by visitors to electronic bulletin board systems but also to compare that language to other types, such as speech or print. The advantage of this approach was that it enabled comparisons: Collot and Belmore could use the same “ruler” to measure language produced in various print and digital settings, which allowed them to argue that the language they encountered online was different from other kinds of language.

Rather than define codes ahead of time, some researchers let codes emerge through a preliminary reading of the material they intend to code. This is a method advocated by Huckin (1992) and described by Barton (2002). Barton described how she pursued a hunch that inexperienced writers use evidentials differently from their experienced counterparts. As Barton wrote, “evidentials are a form of metadiscourse used to express attitudes toward knowledge (e.g., probably, generally, certainly)” (p. 25; author’s emphasis). As a teacher as well as researcher, Barton believed that the way students used evidentials was significantly different from the way experienced writers used them. Her codes emerged from an initial reading of the types of texts she wanted to study; then she devised a method for studying student texts more formally (Barton offers an illuminating reflection on her methods on pp. 36–38).

Just as Barton (2002) described, I learned that a set of codes can emerge from an initial reading of a potential corpus. When I wanted to study the use of Web sites to promote campus–community partnerships, I began with two tasks: I read about ideals for networked computers and civic participation, and I reviewed sample sites. From my reading, I found general consensus regarding ways that networked computers could be used to empower citizens and improve government service, such as creating new forums for deliberation and improving access to elected officials
and information. I also began to notice that the few sites I knew of shared many features with civic Web sites. I reasoned that if campuses were also trying to empower citizens through community outreach, then they should be using networked computers in ways similar to the ideals I was reading about. I defined my own set of codes (based on readings about the ideals of civic computing) and then sampled sites (see the earlier section on defining sets and selecting samples). By defining my own set of codes, I hoped to explore the extent to which campus uses of networked computers met or deviated from ideals expressed in the literature.

Whatever method you choose for defining a set of codes, you need to address several issues. First, all codes in a set must be defined at the same level. That is, the codes should identify the same type of unit. You may decide to code words or exchanges, for example, but not both simultaneously within the same set. You can, of course, code a larger rhetorical unit and then words within that unit. But you would do so with two different sets of codes—a two-tier coding system. You could, for instance, count the numbers of evidentials used in certain rhetorical units. This would require you to develop a set of codes for evidentials and another for rhetorical units. Second, make your codes as discrete as possible. Some researchers claim that you should not be able to apply two codes to the same unit. As Weare and Lin (2000) put it, the codes must be “comprehensive and mutually exclusive” (p. 284). Other researchers understand that this goal is more an ideal than a realistic objective—it can be especially difficult to meet when coding for latent content. Let’s say you were coding statements made by citizens and experts during an online national forum on air quality sponsored by the United States Environmental Protection Agency. Let’s say also that you had created codes to identify the ways people characterize an environmental problem. For example, some may characterize a problem as an issue of justice. If a person writes, “It is no accident when a landfill is built in a community of economically poor minorities,” you might code it justice. Or they may characterize a problem as one of public health. If a person writes, “If the EPA does not enforce air quality standards during construction of this landfill, asthma rates will increase,” you might code it health. Or they may characterize a problem as an environmental issue. If a person writes, “The failure of the EPA to enforce air quality standards during landfill construction will damage our environment,” you might code it environment. The challenge of coding for latent content is most clearly illustrated in the statement coded environment. What does the person mean by our environment? The person may mean the environment (nature) in general, or she may be referring to the neighborhood where she lives, in which case she may be thinking more about economic or public health issues
than environmental issues. So, should the statement be coded as a public health characterization or as an environmental characterization? In such a case, you must look at the context to understand what the person might mean. Or you may need to define a new code (say, unclear or multiple) to show places where you know that a person is characterizing the problem but the language makes the exact nature of that characterization unclear.

Coding the Data

Once you are ready to begin coding the data, do three things. First, perform at least one test run on a subset of your sample. Do this to determine whether your codes are indeed mutually exclusive—or at least approach that ideal. Also ask yourself whether you are encountering interesting phenomena that the codes do not cover. If you modify your codes significantly, you may even want to run a second test. Second, consider using a spreadsheet program to record your codes, as is illustrated in Table 10.1. Record the name of the file in which the coded passage appears, the code you gave to that passage, the passage itself, and perhaps your comments on the passage or on why you coded it as you did. By using spreadsheet software, you can easily count codes when the time comes. (For instance, in Microsoft Excel, you can use the AutoFilter feature to display one type of code at a time.)

Third, and perhaps most important, keep a journal. Reflect on your progress and note potential problems or interesting dilemmas that arise. Note for yourself how you resolved those dilemmas. This can be especially useful when coding for latent content. In the previous section, for example, I mentioned the difficulty of coding for latent content when people use the word “environment.” If you decided to give a public health code (rather

<table>
<thead>
<tr>
<th>FILE</th>
<th>CODE</th>
<th>PASSAGE</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.htm</td>
<td>justice</td>
<td>As a predominately minority area, we felt that someone has to stand up against such prejudice.</td>
<td>Landfill proposal characterized as a form of prejudice. See Mirel (1994) on sources of outrage in risk communication.</td>
</tr>
<tr>
<td>2.htm</td>
<td>health</td>
<td>Citizens in the Midwest may soon face a serious health threat.</td>
<td>The writer defines the landfill as a public health issue.</td>
</tr>
</tbody>
</table>
than an environment code) to statements such as “The failure of the EPA to enforce air quality standards damages our environment” whenever they follow statements about a group of people, then you ought to note that decision.

Testing Reliability

With data coding, reliability is established by showing that different people would code a set of texts, visuals, audio, video, and so forth similarly. Reliability is usually determined by training others to code at least a subset of the phenomena (approximately 10% to 20%) that the researcher coded. The more that other coders and the researcher agree, the higher the rate of reliability. Whether you would want to test for reliability depends on at least two factors. First, it depends on whether you are coding manifest content or latent content. If you are coding manifest content and are taking a more quantitative approach, then you should be able to measure reliability. If, however, you are coding latent content and are taking a more qualitative approach, then it becomes more difficult to measure reliability because of the degree of interpretation. Second, on a practical level, whether to test for reliability also depends on the audience for your research. As McMillan (2000) noted, “an important requirement for any social-science based research is that it be carried out in such a way that its results can be verified by other investigators” (p. 88). Many readers will expect evidence of interrater reliability, especially readers of journals that publish more traditional forms of social scientific research. Other readers, however, don’t expect an account of reliability. In the journal Computers and Composition, for example, you’ll regularly find studies that involved data coding but seldom find reports of interrater reliability. One way to assure adequate reliability is to conduct pre- and post-tests. After you test codes initially yourself, but before you code your entire sample, ask one or two raters to test the codes as well. Use their feedback to help you refine your codes. Once you have coded your entire sample, ask the raters to code another subset of that sample. Use the post-test as the basis for measuring reliability.

Simple agreement can be measured by dividing the number of agreements between two coders by the number of coding decisions they made. For example, if you and your rater coded 20 items in one text, and if the two of you agreed on 17 of them, then you would divide 17 by 20. Your simple agreement would be 85%. Although useful as a basic measure, such agreement does not account for chance. It is possible, of course, that two
people agreed on some percentage of coded units simply by accident. You can minimize the effects of chance by using Cohen's Kappa; Geisler (2004) described this formula (see pp. 79-84) as did Janice Lauer and William Asher (1988, see pp. 261-263). Cohen's Kappa can be calculated by hand or automatically by using a calculator designed for the purpose. When I calculated reliability for my study of campus-community partnership Web sites, I used a convenient online calculator.

As you plan to test your coding, remember that some types of digital writing are modified regularly, especially Web pages. You may want to archive a set of Web pages to be coded, thus ensuring that other coders see the same set of pages. In my study of campus-community partnerships, for example, I burned a CD of Web pages for each coder. Remember, also, that many hypertexts are designed to be explored in numerous ways. Give your raters instructions on how to work through the pages. If you are saving an archive of pages, you could rename each file to indicate the order in which they should be read (e.g., 01.htm, 02.htm, 03.htm). Ask raters to read each file in order and to ignore links within any page. Such a precaution will save you time because, with any luck, your sequence of codes and your rater's sequence of codes will appear in about the same order. You won't have to spend time sorting your rater's sequence so that you can compare it to yours.

Analyzing Results

Essentially, analysis involves finding patterns and interesting anomalies in the coded data. To some extent, analysis may be an intuitive process based upon a researcher's ability to find interesting trends in the data. It can be approached, however, from a more quantitative point of view. Geisler (2004), in Analyzing Streams of Language: Twelve Steps to the Systematic Coding of Text, Talk, and Other Verbal Data, devoted several chapters to ways of detecting patterns, including frequency (which measures how often each code was used) and distribution (which measures the degree to which codes appeared in one setting or another). When I analyzed campus-community Web sites, for example, I was struck by how often Web resources were used either to make an institution's reports available or to promote an institution. (With one institution that I coded, 83% of the units were devoted to promotion through such texts as testimonials, descriptions of staff, and descriptions of past programs.) I based my analysis, in other words, on frequency. I was struck by the focus on sharing reports (essentially a one-way form of communication) and promoting institutions, because those
moves were preferred so heavily over other communicative acts, such as online collaboration and dialogue. What I was seeing differed significantly from ideals that I held regarding the use of networked computers for campus-community collaboration.

FROM METHOD TO METHODOLOGY

You may well ask whether data coding merits the effort—merits all that defining of codes and units and all that sorting, counting, and analysis. What might it enable that other methods might not? Wouldn’t a simple “close reading” of texts be enough? As with most issues, the best answer is that it depends. It depends on your assumptions regarding the epistemological status of texts, on whether the method can support a commitment to others in research projects, and on whether you view research as reflective practice.

Data Coding and the Epistemological Status of Texts

Whether data coding is worth your time depends in part on what you assume about texts and what we may learn from them. I believe that texts in general, and digital texts in particular, are worth analyzing because they reveal important characteristics of culture and human behavior. Their study can also help students work more reflectively with digital media. Some postmodern critics would say that texts cannot be studied as sites “of determinable meaning from which concrete, systematic cultural expectations can appropriately emerge” (Thomas, 1994, n.p.). If texts are indeterminate, then any inference stemming from them is merely the product of interpretation; the results of data coding tells us more about the person doing the coding than about the texts being coded. Although I grant that texts are open to interpretation, I reject the argument that most texts are indeterminate to a degree that renders data coding useless. As with most philosophical issues, truth lies somewhere in the middle. Consider a simple set of directions: With a few well-worded phrases (e.g., “turn left at the light and proceed for 1 mile”), most people can find a place they’ve not visited before. Or consider written policy: Texts that contain criteria (e.g., “students must have at least a 3.0 GPA to be eligible to interview on campus”)
can have a profound effect on social patterns. A passage may indeed be interpreted in several ways, but it usually cannot be interpreted in any way (unless perhaps we're talking about something like Surrealist poetry). Data coding is not random interpretation disguised as social science.

But, some might say that data coding allows only indirect, inferential evidence of culture and human behavior. It would thus be better to study such behavior directly. We should interview participants and observe group meetings, for example, rather than code Web sites for evidence of how digital writing is used in campus–community partnerships. I accept the argument that other forms of observation and analysis are important. Huckin (2004) rightly claimed that data coding (he was writing specifically about content analysis) should be combined with "more sophisticated forms of discourse analysis" (p. 13). However, I am skeptical of the argument that observation of human behavior is more direct and therefore less susceptible to interpretation. As Thomas (1994) argued, all behaviors essentially create "texts" that must be interpreted (n.p.). If I want to study the assumptions that go into the development of campus–community partnerships, I could observe and interview a group of people as they develop such a partnership. I could videotape meetings to record gestures, tone of voice, facial features, and other variables. Still, I would have to interpret those moves. I would be left, in other words, with a new set of texts to be interpreted.

Rather than privilege the use of interviews and observations as somehow more honest, direct, or unfiltered than forms of text analysis, we should acknowledge what each method of data gathering is likely to reveal. In the study of digital writing, data coding can tell us about how the Internet and multimedia are being used, about how group and individual identities are constituted and maintained online, about people's attitudes toward the new media, and more. Data coding, in other words, helps us answer many who and what questions, such as Who is participating in this group? What are they saying? For instance, we can explore the personal and professional identities some professional women and men create in part through the photographs and drawings they include on their home pages (Hawisher & Sullivan, 1999; Hess, 2002). Data coding is also uniquely qualified to address questions of how much? Sometimes it simply helps to know how often something occurs. How often, for example, do Japanese teachers of English use social relation markers to indicate hierarchical or egalitarian power relations (Matsuda, 2002)? How often are blacks presented photographically in positive or negative social activities (Smith & Price, 2005)?

When we analyze texts, we analyze products. What we cannot analyze are processes. In other words, data coding may help us explain what is pre-
sented in texts, and how often, but it does not help us understand why the
texts look like they do. As Berkenkotter (2002a) pointed out, textual analy-
sis has “been criticized as being too narrow and reductive... as it tends to
focus on textual phenomenon per se, and not to attend to context” (p. 49).
That would be true if a researcher were to rely on data coding alone.
However, textual analytic methods can be combined with others, such as
interviews and participant observation, to provide such context.

Data Coding and Commitment to Others

Because data coding focuses on patterns and trends in texts rather than on
individuals, it is fair to ask whether the method has any place within criti-
cal research. Sullivan and Porter (1997) assumed that critical research
begins and ends with a commitment to research participants, which is its
“ultimate aim” (pp. ix–x). This commitment led Sullivan and Porter to
define overlapping goals for critical research:

- Respect difference. Sullivan and Porter argued that researchers
  must recognize differences between groups and individuals, and
  they must avoid research practices that ignore those differences.
  Moreover, researchers also should not “instrumentalize” people—
  that is, should not use others for their own ends, such as publica-
  tion.
- Care for others. Sullivan and Porter argued that researchers
  “should not proceed primarily out of a motivation to discover
  new knowledge, but rather should be motivated by a commit-
  ment to the participants, a concern for their welfare” (p. 113).
  This care should be based not on some universal commitment
  but on a local commitment to the people with whom a researcher
  works.
- Promote access to rhetorical procedures enabling justice. “Justice is
  realizable,” wrote Sullivan and Porter, “only when people have
  access to the mechanisms of policy and decisionmaking” (p.
  115). People should have the ability to influence situations that
  affect them, including research situations. Research participants
  should have “a say in determining their status within research
  projects” (p. 118).
- Liberate the oppressed through empowerment of participants. With
  this goal, Sullivan and Porter rejected the notion that a researcher
  is some sort of powerful liberator. Rather, they pointed out that
"researchers and participants" should enjoy "a reciprocal relationship" in which participants as well as "our research communities" benefit from research (p. 124). Participants should have a say in how research is presented, and they should be able to use the results of that research in ways that suit them.

Given that critical research begins with a commitment to others, its focus is directed toward action rather than observation. Critical research is about working with others in order to improve real conditions. When applied to digital writing, critical research obliges researchers to "pay attention to users in context" (Sullivan & Porter, p. 107). Data coding, however, does not even require a research participant. What role, if any, might the method play if a commitment to others is a cornerstone to methodology?

If commitment to others forms the core of your methodology, you may simply have to acknowledge that data coding may be most vulnerable to critique. Coding offers no direct support for research participants; however, it may in some situations contribute to the ethical treatment of research participants, and it could be a form of research in which participants themselves could engage. For example, if I were working with a group of local environmental activists to try to help them create a Web site and database, then I would want to describe for them how others have created such resources (or, even better, they may ask me to help them create their own data-coding study). I would want to reveal patterns to help them more fully articulate their own approach to the problem. Or consider the value of data coding for participants in Heidi McKee's (2002) study of the dynamics of interracial electronic communication. If participants could see patterns in the ways they talk about race, in the assumptions revealed by their postings, could such self-awareness improve interracial communication in that group? Or, what if the students completed a data-coding exercise documenting interracial communication in another public forum? Could the knowledge gathered from such a study help students as they began their own forum? Although data coding may not have the direct interpersonal appeal of, say, participant observation, it could be used as a lens to help participants see how others communicate or as a mirror to reveal for participants their own patterns of communication.

Data Coding and Reflection

Although data coding may offer only indirect commitment to others, it may have its greatest value as a form of reflective practice. Sullivan and Porter
(1997) argued that critical researchers should recognize "the rhetorical nature of research activity" (p. ix). They should

recognize that such issues as the posing of research questions, the framing of a research study, and the choice of methods are discursive activities. Because they are discursive, researchers must keep themselves alert to those elements in their practices that adopt positions or attitudes or actions without reflection. (p. 16)

The need for reflection always exists because method influences results. As Grant-Davie (1992) wrote, "what researchers find in the data is influenced by what they look for" (p. 273); I would say results are also influenced by how they look. We know, moreover, that researchers must devise and revise methods as they go. Few would argue that researchers should pick a pre-existing method and follow it lock step. Researchers should interrogate their own work; if a researcher must analyze a set of texts, then data coding actually provides a mechanism for reflecting on such analysis.

Conceiving of research as a journey from theoretical assumption to observation to analysis, Titscher, Meyer, Wodak, and Vetter (2000) wrote:

If one proceeds systematically wrong turnings are avoidable. Methodical procedure can, like Ariadne’s thread, guarantee the researcher a safe route back. By giving them experience along the way, methodical procedure may also assist those investigators who look over their shoulders and see their starting point differently, even deciding not to go back what to find other more interesting starting points. No matter how the investigative journey may turn out, methodical procedure will make it easier to report findings and to compile reports of experience. (p. 6)

Data coding is most valuable, in my view, because it asks researchers to create a record of decisions made during a project. Researchers must record their method for finding texts (which is especially important in studies of the Web where no comprehensive index exists to aid selection); their rationale for choosing texts they find; and the concepts they used to analyze words, visuals, audio, video, and links. With this record, a researcher can reflect more fully on decisions made along the way.

Coding is also valuable because the record a researcher makes is then available for others as they examine the basis for a researcher’s analysis. As Thomas (1994) wrote, methods of data coding (such as content analysis) are not "objective" in the sense that they present a "socially uncontaminated" path to Truth. Rather,
what makes content analysis "objective" is that, as much as possible, the researcher is obliged to make public the bases for the sampling and analytic choices. Most important, the content analyst's culminating interpretations are tied to these revealed procedures. (n.p.)

Granted, Barton (2002) reminded us that the methods sections in which our research process is revealed are often kept short and vague, that sometimes they are even contradictory. Despite the limitations of some methods sections, data coding remains valuable, not because it confers a kind of social scientific legitimacy, but because it enables the kind of complete disclosure and reflection that Sullivan and Porter (1997) advocated.

CONCLUSION

Data coding presents researchers with a trade-off. As Grant-Davie (1992) noted, researchers "sacrifice detail" when they code, but "they do so in order to translate the data into more abstract forms, forms that can reveal patterns that would be lost in the mass and complexity of uncoded data" (p. 284). Data coding is about finding patterns, about seeing big pictures in large amounts of data. In the search for large-scale trends, though, individuals can get lost. The key to data coding, then, is knowing what it will reveal and conceal, and to combine it with other methods in order to create a more complete picture.

NOTES


2. Although I could have focused on one approach, such as content analysis, I have chosen instead to focus on data coding—on the steps that such social scientific methods share. As Titscher et al. (2000) argued, the boundaries between methods such as content or discourse analysis blur to the point of near unintelligibility. I do not mean to criticize when I describe the difficulty of distinguishing one form of text analysis from another. I believe, however, that the method of selecting a corpus, coding data, and revealing one's method can be discussed in general.
3. To gather your own copies for coding is considered fair use, but you should destroy the copies once you have completed the project. If you publish your work and decide to quote extensively from a copy, especially if you want to include a screen capture, then most likely you will need permission from the copyright holder.

4. For a more comprehensive list of such units, see chapter 3 in Geisler (2004).

5. Smith and Price (2005) found that blacks were significantly underrepresented compared to whites in nondailies but that when blacks were depicted, it was in much more positive ways than in daily papers.

6. Several software packages have been developed to support data coding, including Atlas, Ethnograph, Hyperresearch, and NVIVO. I have used Microsoft Excel to code data because I'm already familiar with it and I already own it. That is, it presents no learning curve and costs me nothing extra.