

# AN APPROACH TO THE MANAGEMENT AND CODING OF QUALITATIVE DATA USING MICROCOMPUTERS

JOEL GITTELSON

Social scientists are increasingly asked to contribute to action-research projects using their skills as qualitative researchers. The large volume of textual data produced by even a fairly small qualitative study poses special problems in terms of data management. Recent advances in the use of microcomputers for qualitative data management and analysis provide at least a partial solution. The paper describes an effective method for managing qualitative data on microcomputers using textual data management software and incorporating appropriate coding procedures.

Dr. Joel Gittelsohn (Ph.D.) is working in the Division of Human Nutrition, Department of International Health, School of Hygiene and Public Health, The Johns Hopkins University, 615 N. Wolfe St. Baltimore, MD 21205.

## Introduction

Social scientists participate in action-research projects in various capacities: as an operations researcher in an urban teen pregnancy project, an ethnographer in an AIDs intervention, as a social worker for a weaning intervention, and so on. In recent years, there has been a rapid growth in the use of qualitative or semi-qualitative data-gathering in connection with health and nutrition programmes. The insights provided by qualitative data have proven valuable for programme planning, implementation and evaluation (Coreil et al., 1989; Bentley et al., 1990). Researchers using qualitative methods generate large amounts of context-rich descriptive textual data (Peshkin, 1988). For instance, social workers may develop hundreds of detailed individual case studies. Commonly, the contribution of the social scientist in these circumstances has been as an advisor and as a producer of reports. Commonly, in these reports, only the barest summaries of the qualitative data collected are presented, emphasizing numbers and counts, the quantitative aspects of qualitative data. Qualitative data presented in numeric form (usually as tabulated lists) has minimal influence on decision-making at the policy and planning levels. It is often viewed as poor quantitative research, based on an inadequate, unrepresentative sample. There are at least three ways to deal with this perception and to improve the utilisation of qualitative data in action-research projects: (1) improve data collection; (2) improve data management systems; and (3) develop new means of qualitative data analysis (Looker et al., 1989). This paper focuses on the second component: improving qualitative data management.

This paper describes an approach for managing and coding qualitative (textual) data using microcomputers. The use of microcomputers in quantitative research is not new, but their use in qualitative research is still in its formative stages (Richards and Richards, 1987; Fielding and Lee, 1991). The method presented here is especially useful for working with qualitative data gathered as part of a multidisciplinary research/intervention project. The method is currently being used to manage textual data collected as part of a teenage pregnancy intervention project in Baltimore, for dietary management of diarrhoea research in Nigeria, and for NGOs working in women's reproductive health in India and Nepal (Bentley et al., 1990; Bentley et al., 1991). Besides these types of interdisciplinary projects, I believe that this method is also effective for individual social scientists conducting more traditional ethnographic fieldwork, and for use with the intensive case

study data generated by social workers. It is my contention that an organised, accessible database of qualitative data will allow social scientists to increase their contribution to applied intervention projects, as well as to more theoretically-oriented studies.

### **The Challenge of Textual Data**

Qualitative research commonly generates thousands of pages of written (or "expanded") notes. Within these notes lie interviews with key informants, transcripts of focus group sessions, observations of behaviours and events, secondary data (such as clinic records and genealogy charts), and the social scientist's comments and thoughts about what these data all mean (Bernard, 1986). These diverse data present a tremendous challenge for management and analysis. How should these notes be organised? Should they be coded? If so, in what system? How can the investigator locate examples within the large body of textual data that will allow him or her to confirm or reject their theoretical assumptions?

Bernard (1988) writes, "No single researcher, working alone for less than two years, can produce more field notes than she or he can grasp by pawing and shuffling through them." This may be true for the lone ethnographer or social worker, but as stated earlier, increasingly, qualitative data are being gathered by teams of data collectors. If four data collectors (an investigator plus three research assistants), produce just eight pages of notes each day for a (relatively) short three month study, the result is over 2500 pages of fieldnotes! Not only is this a large number to "paw and shuffle" through, but the investigator may not be entirely familiar with the data collected by his or her assistants.

### **Text Management Using Microcomputers**

The solution proposed here for the managing of textual data is the use of a word processing programme (such as Word Star or Word Perfect) and a text retrieval and management programme (such as ZyINDEX). Note that this solution requires that all textual data be entered on the microcomputer. While this may constitute a difficult step for many individuals/organisations, the results in terms of accessibility of data make the effort worth-while.

There are dozens of word processing programmes available. In the USA, Word Perfect is the most commonly used programme — while in India, Word Star holds away. When selecting a word processing programme for entry of textual data, the following features are essential: spell checking, word searching, user-definable dictionary and some kind of macro-key function (a type of small user-develop programme, where multiple keystrokes can be accessed with a single Keystroke). The uses of these special word processing features will be described later in this article.

A complete approach for managing, coding and retrieving qualitative data is outlined in Figure 1. In the next three paragraphs I present an overview of the process of managing textual data on microcomputers, then cover specific topics in greater detail.

"Raw" fieldnotes or case history notes (i.e., handwritten "jottings" or notes made during the actual data collection) should be typed as expanded fieldnotes on a microcomputer. Alternatively, expanded fieldnotes can be written on paper and given to a professional word processor for entry. These expanded fieldnotes should then be "spell-checked" for errors<sup>2</sup>. In order to use this method, interviews will have to be translated into English. Key concepts in the local language

Figure 1

**MANAGING QUALITATIVE DATA USING A MICROCOMPUTER SYSTEM**

<i>Process</i>	<i>Tool Used</i>
1. Raw Field Notes	Hand-Written/Tape Recorded
2. Expanded Field Notes	Either expanded into notebooks and later entered into Word Perfect; or expanded directly into Word Perfect
3. Spell-Checking	Word Star or Word Perfect
4. Review of Expanded Field Notes & Derivation of Preliminary Codes	Word Star or Word Perfect and Word Perfect Macros
5. Coding of Expanded Field Notes	Word Star or Word Perfect Macros
6. Indexing Expanded Field Notes and Codes	ZyINDEX
7. Searching Through Expanded Field Notes and Codes	ZyINDEX

(such as illness terms) can be entered in a romanised format<sup>3</sup>. Careful attention should be paid to keeping the spelling of local terms consistent.

I suggest that each data file have a unique, meaningful filename and each file represent a discrete data collection unit: i.e., one interview, one observation of a key event, one focus group session. In projects I have worked on, we use a format for naming files where the first two characters indicate the type of data collected (i.e., key informant interview, focus groups); the next six indicate the date of data collection (in year, month, day order). The first two characters of the filename extension are the initials of the data collector and the last digit is an order number for the day (for example, KI900711.JG1 would be the filename for the first key informant interview conducted by Joel Gittelsohn on July 11, 1990<sup>4</sup>).

These expanded notes should be reviewed, and a preliminary coding list derived. The notes should then be coded within the word processing programme. Once the preliminary coding process is complete, the notes can be indexed using ZyINDEX. Preliminary searches of the data should be conducted at this point. If the investigator is still in the field, these preliminary searches can help him or her identify "gaps" in the data (i.e. areas for continued research). The coding list should be refined and expanded at this point as well.

### **How Text Management Programmes (like ZyINDEX) Work**

ZyINDEX offers "full-text" information retrieval, that is, finding any word, phrase, or number or combination of these elements wherever they occur in all your documents. ZyINDEX's first program

function-indexing, works by creating a list of *every word and number in* specified text files, keeping track of what file (or files) each word is in and where the word is located within each file. Common words like articles and conjunctions (e.g. the, or, and ...), called "noise words", are ignored to save space on the computer's hard disk. ZyINDEX manages existing data files without altering them and requires no specially identified key-words. ZyINDEX can recognise a wide variety of word processing formats (such as Word Star and Word Perfect), plus LOTUS 1-2-3, dBASE III, and ASCII files.

ZyINDEX second programme function, searching, uses the specially created index lists to locate user-specified text "strings" (groups of words that the searcher desires to locate). Searches may be refined by the use of Boolean operators (AND, OR, NOT), wildcard characters, parentheses and a "search within" feature which lets you specify the proximity (in number of words) of two search items. Searching in ZyINDEX is very fast. Depending on the speed of your microcomputer, the program can search through thousands of pages of text in just a few seconds.

Once located using the search feature, important text can be marked and extracted for analysis. Unlike some programs which pull out isolated "hits", ZyINDEX permits the searcher to view the full text, allowing the user to see the context of the "hits".

### Coding Qualitative Data

What are codes? A code is an "abbreviation or symbol applied to a segment of words...in order to *classify* the words...they usually derive from research questions, hypotheses, key concepts or important themes" (Miles and Huberman, 1984), Codes have a number of different functions relating to textual data, including organisation, retrieval of information, assembly and reduction of data into analysable units. As well, Miles and Huberman (1984) observe that "coding is a form of continuing analysis...it sets the agenda for the next wave of data collection".

There are many different systems for coding qualitative data. These can be divided into three basic categories: numeric, mnemonic and word-based systems. In any coding system, codes represent key concepts and ideas in the text. For instance, a code may signify: **"here is an example of the expression of female power within the household"**. It is very time-consuming to write that whole sentence down manually on your notes each time it is relevant to do so, therefore, abbreviated forms (i.e., codes) are used. Numeric codes, where number represent key concepts and ideas, are usually the most concise form of coding. It is far easier to write a number code (i.e., 562.02) than it is to write out the whole phrase above. On the other hand, numbers are hard to remember. It is easy for a coder to miscode textual data when numeric coding systems are utilised. Mnemonic coding systems, where a set of letters are used to represent the key concept or idea, are a popular alternative to numeric coding. For instance, using a mnemonic system, the phrase above could be coded as: FEM\_\_POW\_\_HH.EXM. This kind of code makes more sense to the coder, is easier to use, and may even be understandable to the lay user. Finally, word codes use whole words to indicate key concepts and ideas in the text, such as EXAMPLE OF FEMALE POWER IN HOUSEHOLD. Generally, while whole word codes are much clearer to the user (particularly to the lay user) they have not been widely adopted because it takes too long to code qualitative data manually in this manner.

In addition to the *type* of code to apply, codes can be applied to data at several *levels of abstraction*. Miles and Huberman (1984) describe three coding levels, descriptive, interpretive, and explanatory. At the lowest level are codes that describe things as they are found in the dataset

(descriptive), at a higher level codes represent judgements by the coder about what is happening in the data (interpretive), and at the highest level they relate important concepts according to inferred patterns (explanatory), "pattern codes are explanatory or inferential codes, ones that identify an emergent theme, pattern, or explanation that the site suggests to the analyst" (Miles and Huberman, 1984).

A third aspect of coding relates to the number of different codes to apply to a textual dataset. The intensive coding schemes described by several authors (Bernard, 1988) have several deficiencies:

- (1) Such coding schemes must be constantly updated by modification and/or additions to accommodate the needs of a dynamic database, one that grows and changes as the needs and purposes of the database change.
- (2) Intensive coding schemes are not "user-friendly". Many of these systems involve hundreds of numeric or mnemonic codes which require a great deal of time and sophistication to learn and apply correctly.
- (3) As professional/trained coders are required to learn and apply the multitude of complex codes, increased expenses are entailed.
- (4) The "accessibility" of large qualitative datasets is very low when complex coding systems (particularly numeric or mnemonic) are utilised, especially for lay users.

### **Coding for a Microcomputer-Based Text Management System**

Use of microcomputer-based textual management system requires that text be entered using a word processing programme (such as Word Star or Word Perfect). However, once textual data files have been entered on a word processing programme and indexed using textual data management software (such as, ZyINDEX), it can still be difficult to find certain kinds of information. For instance, a direct observation of a woman ordering her husband to buy her a dress might be a good example of "household decision-making" in action, but how would you locate it? In ZyINDEX every word in the file becomes a "searchable" word, but it is rare that abstract concepts will consistently be part of the data to be located. The solution lies in the use of codes.

Developing a coding scheme can be a useful analytic and retrieval tool, but no coding scheme can cover every possible way of looking at textual data without being impossibly large, unwieldy and complex. I propose that while the coding of qualitative data is necessary, it should be minimised when using text management and retrieval software like ZyINDEX.

If coding of qualitative data is kept to the minimum, the result is a more user-friendly and accessible database. A reduction in the number of codes improves intra- and inter-coder reliability. As well, there can be a tremendous saving in time and money. The coding method described below allows non-social scientists to maintain and access a textual database.

Components of a "Minimalist" Coding System for Microcomputers:

1. Use a word-based coding system instead of numbers or mnemonics to reduce confusion. Word codes are more likely to be understood by more people.
2. Do not use words that already appear in the text, as they can be easily searched by text management software, such as ZyINDEX. Thus, lower-level "descriptive" codes are usually not necessary.
3. Codes should be used for concepts at high and occasionally medium levels of abstraction (explanatory and interpretive).

4. While only a rough guideline, a good minimalist coding system should have less than 50 codes. Expansion of the system may be required as the research grows to cover new areas of investigation.
5. If required, a word-based coding system can be tied into other coding systems. For instance, a word-based system tied into Murdock's (1971) Outline of Cultural Materials coding system might yield the following hybrid code:  
 [#STATUS OF WOMEN—562.02#]

Figure 2 gives an example of a coding system developed by an Indian NGO for use with a qualitative dataset on women's health.

**Figure 2**

**SAMPLE TEXT CODING SCHEME DEVELOPED FOR USE ON  
QUALITATIVE DATA ON WOMEN'S HEALTH**

<i>ALT-KEY</i>	<i>WORD CODES (MNEMONIC CODES)</i>
A	[#ALLOPATHIC TREATMENTS—DOCTORS—(AL.TR)#]
B	(#BARRIERS—HEALTH SEEKING BEHAVIOUR—(BA.HSB)#)
C	[#CAUSES—BEHAVIOURS LEADING TO ILLNESS—EMIC—(CA.LI)#]
D	[#DECISION MAKING PROCESS—HEALTH SEEKING BEHAVIOURS—(HSB.O)#]
E	[#EMIC—CAUSES BEHAVIOURS TO AVOID ILLNESS—PREVENTION—(EC.AL)#1
F	[#FOOD PROHIBITIONS, TABOOS—PRESCRIPTIONS, PREFERENCES—DIETS—(FP.DI)#]
H	[#HOME REMEDIES—TREATMENT—(TR.H)#J
I	[#INDIGENOUS—TRADITIONAL—TREATMENT—(IN.TR)#]
L	[#LIFE HISTORY—CASE STUDY—STUDY—(LH.CS)#]
M	[#MOTIVATIONS—HEALTH SEEKING BEHAVIOURS—(HSB.M)#]
O	[#OBSTETRIC HISTORY—BIRTHS—MORTALITY—FERTILITY- (OH.F)#]
R	[#ROLE, STATUS OF WOMEN—POWER—(RS.WP)#]
S	[#SOCIO ECONOMIC—CAUSES—(SE.CA)#J
T	[#TECHNIQUES—METHODOLOGY—DATA COLLECTION—(TE.ME)#]
W	[#WOMEN—GIRL CHANGES—BELIEFS—(WG.CB)#]
X	[#FOLK PHYSIOLOGY—BODY CONCEPT—FUNCTION—(FP.BC)#]
Z	[#TERMINOLOGY—DEFINITION—TAXONOMY—(TE.TAX)#]

### How to Apply Codes to Textual Data Using Microcomputers

All textual data should be entered on some type of word-processing programme. Codes should be clearly marked as separate from the textual data. I recommend making codes appear different from regular text by capitalising them, boldfacing them, putting them in brackets and by adding a symbol on each end of the code (for instance, [#STATUS OF WOMEN#]). The extra symbol allows you to search exclusive or inclusive of the code. For a limited number of codes, it is possible to "hot-key" a number of codes — wherein codes are inserted into the text in response to one or two keystrokes. For instance, in the WordPerfect software programme, macros can be developed using the "alt" key in combination with a letter to print the code. This saves a lot of time in the coding process, and is more rapid than manually entering numeric or mnemonic codes. In terms of location of the codes, I recommend placing the codes immediately following the text they relate to.



3. What information do we have on what people believe are the causes of diarrhoea (middle-level abstraction)?
4. What information do we have on the overall complexity of the local medical system (high-level abstraction)? For instance, do people commonly use more than one type of health provider?

The procedure for searching the indexed database would differ depending on the type of research question asked. As a general rule of thumb, it makes sense to start with the simple term and simple category types of questions. If these basic questions cannot be answered by the materials in the database, it makes no sense to move onto more complex questions. In the following examples, search "requests" are written in the syntax preferred by ZYINDEX.

For a *simple term search*, just use words as they appear in the text. This works especially well with nouns. For example, to find information on the concept "fever", the appropriate search term might look like this:

(fever or joro)

The simple search shown above would retrieve all files that contain either the word fever or the word joro<sup>6</sup>. All the searcher needs to know is the English word and appropriate local word for the term.

For a *simple category search*, use combinations of words (usually synonyms) as they appear in text. This also works especially well with nouns. The format for a category search is: (synonym1 or synonym2 or synonym3 ... or localterm1 or localterm2 or localterm3...). For example, to find information on the concept "diarrhoea," an appropriate search term might look like this:

(diarrhoea or faeces or dysentary or **gu** or disaa)<sup>7</sup>

The searcher here needs to know all the English words and local words for the category. ZYINDEX has several special features (thesaurus, wild cards, access to indexed word lists) to assist the user in determining the available synonyms for a particular key concept or term.

For a *middle-level abstraction search*, again use combinations of synonyms as they appear in text (although codes may be needed as well). This type of search usually takes the form of a set of noun synonyms matched with a set of verb synonyms, usually linked by a proximity term. The format for a mid-level abstraction search is: (synonym1 or synonym2 ... or localterm1 or localterm2...) w/#(synonym1 or synonym2 ... or localterm1 or localterm2). For example, to find information on the middle-level abstraction "causes of diarrhoea," an appropriate search term might look like this:

(diarrhoea or faeces or dysentary or **gu** or **disaa**) w/10 (**cause\*** or **get\*** or **make**)

The search above requests that all diarrhoea-related terms within 10 words of any cause-related term to be retrieved. It is crucial that the searcher needs to know all the English words and local words for the terms/categories and how people talk about them. Especially the types of expressions/verbs which people use when they talk about diarrhoea and its causes.

For a *high-level abstraction search*, it is likely that codes will be required, perhaps in combination with nouns and verbs. For example, to find information on the high-level abstraction "pluralistic illness explanatory system," an appropriate search term might be:

**[#PLURALISTIC ILLNESS EXPLANATION\*] w/10 (diarrhoea or faeces or dysentary or gu or disaa)**

In the preceding search request, any use of the code [#PLURALISTIC ILLNESS EXPLANATION\*] within 10 words of diarrhoea-related term will be retrieved. The searcher needs to be familiar with the coding system as well as all English words and local words for the terms/categories and how people talk about them.

At the very *foundation of the process* for managing textual data described in this paper is careful attention to producing good expanded fieldnotes. This cannot be overemphasised. ZYINDEX makes practically every word in each one of your data files a "key word," so the more attention you pay to getting it all down accurately, the more effective and efficient your searches will be.

## Conclusions

Multi-disciplinary research and intervention projects could make good use of this microcomputer-based method for managing and accessing their qualitative data. A detailed qualitative database is a valuable resource for an organisation. Unlike quantitative data, a qualitative database does not require the assistance of a statistician or programmer to produce useful information. Certainly, care must be exercised that unfounded conclusions are not drawn *on* the basis of inadequate use of the dataset, but this is a problem that is common to quantitative datasets as well.

A qualitative database provides a number of attractive features to organisations, including its great depth and breadth of information, approachability and expandability. As organisations evolve and desire to investigate new areas of concern, they are likely to find that their qualitative database supplies a lot of information on topics they had not expressly investigated before. This is because the qualitative research approach emphasises *contextualisation*. Beliefs and behaviours do not *occur in isolation*, but are part of larger patterns of interlinked social and cultural features. But these patterns are complex and difficult to tease out and describe. A microcomputer-based management system is an effective means of organising and accessing qualitative data, bringing out its full contribution for action-research projects.

## NOTES

1. I have reviewed several other textual management programmes (e.g. Ask Sam, *Word Cruncher*, GOFER, etc.), but have found ZYINDEX optimal for the purposes described. The programme satisfies the majority of criteria laid out by Wellman and Sims (1990) for textual data management software.  
ZYINDEX operates only on IBM and IBM-compatible machines. Users of Macintosh microcomputer system have some software available for textual data management, including NUDIST (Non-numerical Unstructured Data Indexing Searching and Theorizing) and Hypersoft.
2. Most word processing programmes today have a spell-checking feature.
3. While it is not possible at this point in time, it may soon be feasible to use a special *computer "card"* (such as TRANSCRIPT) so that qualitative data can be entered and managed in the local language directly without translation into English.
4. The advantage of this particular file-naming system is that files can be easily sorted by type and date of data collection.
5. Words or combinations of words in indexed files that meet the user's search criteria.
6. The Nepali word for fever.
7. ZYINDEX has a feature which permits you to write macros to allow you to quickly state certain kinds of search terms. For instance, a search for information on breastfeeding might require a search expression like: (BREAST" or BM or DUDH or MOTHER'S MILK)

## REFERENCES

- Bentley, M.E., Stallings, R. and Gittelsohn, J.  
1990 "Guidelines for the Use of Structured Observations in Health Behaviour Intervention Studies," Document prepared for the World Health Organization/CDD Programme.
- Bernard, H.R.  
1986 "The Construction of Primary Data in Cultural Anthropology," *Current Anthropology*, Vol. 27(4): 382-396.
- 1988 *Research Methods in Cultural Anthropology*, Newbury Park, CA: Sage.
- Coreil, J., Augustin, A., Holt, E. and Halsey, N.A.  
1989 "Use of Ethnographic Research for Instrument Development in a Case-Control Study of Immunization Use in Haiti," *International Journal of Epidemiology*, Vol. 18(4): s33-s37 (Supplement 2).
- Fielding, N.G. and Lee, R.M.  
1991 *Using Computers in Qualitative Research*, Sage Publications: Newbury Park.
- Gittelsohn, J. and Pelto, P.J.  
1992 "Suggestions for the Appropriate Use of Qualitative Research for Developing Health Intervention Programs," (in preparation).
- Jick, T.D.  
1979 "Mixing Qualitative and Quantitative Methods: Triangulation in Action," *Administrative Science Quarterly*, Vol. 24: 602-611.
- Looker, E.D., Denton, M.A. and Davis, C.K.  
1989 "Bridging the Gap: Incorporating Qualitative Data Into Quantitative Analyses," *Social Science Research*, Vol. 18:313-330.
- Miles, M.B. and Huberman, A.M.  
1984 *Qualitative Data Analysis*, Sage Publications: Newbury Park.
- Murdock, G.P.  
1971 *Outline of Cultural Materials* (4th rev. ed., 5th printing with modifications). New Haven, CT: Human Relations Area Files.
- Peshkin, A.  
1988 "Understanding Complexity: A Gift of Qualitative Inquiry," *Anthropology and Education Quarterly*, Vol.19: 415-424.
- Richards, L. and Richards, T.  
1987 "Qualitative Data Analysis: Can Computers Do it?," *ANZJS*, Vol. 23 (1); 23-35.
- Spradley, J.P.  
1979 *The Ethnographic Interview*, Holt, Rinehart and Winston, Inc.: Chicago.
- Stone, L. and Campbell, J.G.  
1984 "The Use and Misuse of Surveys in International Development: An Experiment from Nepal," *Human Organization*, Vol.43 (1): 27-37.
- Weller, S.C. and Romney, A.K.  
1988 *Systematic Data Collection*, Qualitative Research Methods Series No. 10, Sage Publications: Newbury Park.
- Wellman, B. and Sim, S.  
1990 Integrating Textual and Statistical Methods in the Social Sciences, *CAM Newsletter*, Vol. 2 No.1