



Published in final edited form as:

*Qual Sociol.* 1994 ; 17(3): 283–292. doi:10.1007/BF02422256.

## Qualitative Data Analysis with Hypertext: A Case of New York City Crack Dealers

Ali Manwar<sup>1</sup>, Bruce D. Johnson, and Eloise Dunlap

### Abstract

Ethnography has become a useful method in procuring sensitive information from the ‘hidden population’ who may not be accessed with quantitative survey techniques. Researchers are generating huge amounts of qualitative/textual data. Qualitative data require careful planning in storage, coding, retrieval, and analysis. Personal computers have solved data management problems, but data analysis remains problematic. The paper describes some qualitative data management and analytic problems faced by a team of ethnographers engaged in a longitudinal epidemiological study of cocaine and crack distribution/abuse in New York City. Ethnographic data was collected through multi-session open-ended interviews with more than one hundred cocaine/crack dealers and extensive field-notes were kept. Compared to other programs, a hypertext software — Folio Views — was more useful in solving (a) data management and (b) analytical problems. Authors used this software to handle more than twenty-five thousand pages of texts; search and sort the database by any words or codes; and retrieve relevant textual materials needed to complete comparative and thematic analysis. Authors analyzed the data from outsiders' point of view (etic) as well as from the viewpoint of the subject populations (emic).

### Keywords

ethnography; hypertext; drug abuse; crack

### Introduction

In contemporary social research, unlike traditional anthropological studies, ethnography is used to collect data in industrial societies because of its success in collecting information from so called ‘hard to reach’ populations such as drug abusers and those at risk for HIV/AIDS (Kotarba 1990; Adler 1985). Simultaneously, contemporary ethnographic research has ushered in an element of skepticism (Hymes 1976) and a debate about the methodology (Geertz 1973; Marcus and Cushman 1982). Furthermore, textual data, ethnography in its rudimentary form, have presented a challenge for information and social scientists alike. The problem of data management becomes especially complex when volume of data is large. This paper discusses two kinds of issues about textual data: methodological and analytic. It is based on the experience of a team of ethnographers engaged in a longitudinal study of crack cocaine distribution/abuse in New York City.

The Natural History of Crack Distribution/Abuse project has been operational for more than four years. Three ethnographers have been working in this project conducting field work, writing field notes, conducting intensive interviews, and compiling relevant data about crack cocaine use, abuse, and dealing activities in New York City. Street contacts with over 300

---

<sup>1</sup>Direct correspondence to Ali Manwar, Ph.D., National Development and Research Institutes, Inc., 11 Beach Street, New York, NY 10013.

crack distributors and abusers are recorded in field notes, and from this group 80 distributors were interviewed intensively about their life-history and drug distribution activities.

## Nature of Data

In this project we have gathered two kinds of data: interviews and fieldnotes. The fieldnotes contain description of drug-dealing locales, neighborhoods, dealing practices, household operations — norms and behavior, drug market mechanisms, and friends and family-kin networks. Ethnographers wrote extensive field notes describing their observations and relevant information they received through informal conversations with the drug dealers and users.

For interviews, ethnographers followed a pre-planned guide, with the focus on the life history and distribution activities. After obtaining the subject's informed consent, these in-depth interviews were taped and later transcribed verbatim. In the process we have generated more than twenty-five thousand standard pages of textual data. This massive data contains what most ethnographers aspire to achieve — a collection of dialogical discourse between ethnographers and subjects.<sup>1</sup> While fieldnotes provide ethnographers' point of view, interview transcripts furnished lengthy responses of interviewees in their own 'language'. As a result, both emic and etic perspectives (Harris 1968) are available for analysis.

## Open-ended Interviews

Although some structure was provided for conducting the in-depth interviews, it served as a guide for discussion topics rather than a survey protocol. All interviews were open-ended and conducted over several sessions. Participants were free to talk at length on any particular issue they thought significant. Ethnographers guided the discussions by introducing new topics and/or probing topics raised by the respondent. For example, if the interviewer asked them to describe their childhood, the subjects often started with family members and their relationship with parents and siblings then brought up drug use/abuse by family members. In such situations, the interviewer probed the respondent to elaborate upon the modes, frequency, and types of drugs used by family members. Thus, in response to one question several different topics were discussed. Sometimes subjects gave a yes or no answer; the substantive information was in the question rather than in the answer.

## Articulation Ability

The quality of respondents' answers depended on how articulate subjects were, the sharpness of recollection, and mood at the time of the interview. Although the subjects were generally sober (i.e., not 'high' on drugs or alcohol) during the interviews, responses were always in their idiosyncratic street language. Their answers were shrouded in 'emic' expressions and terminology, which assumed that the listener could infer their unstated meanings. In decoding these information, we needed to be knowledgeable about street-argot and terminologies, their symbols, and the meaning of the symbols. The problems of analysts and interpretations are discussed below.

We faced a number of problems in handling the huge amount of data that we have collected. Most of these problems were overcome with a new hypertext software called Folio Views (Folio Corporation, Provo, Utah). This new hypertext package very nearly meets all the requirements for easy handling of textual data. It can convert to and from most standard word processing programs, reorganize text files, and retrieve and compile text rapidly. It has the

---

<sup>1</sup>According to Marcus and Cushman: "The dialogical model depends on a representation of the actual discourse of fieldwork, and while no less a construction of the ethnographic writer than Geertz's textualization, it at least attempts to stay close in its representation of data to the material from which cultural texts are abstracted for interpretation" (1982:43).

capacity to link related information regardless of their location in the database(s). The search commands are easy to learn, with results seen instantaneously on the monitor.

One useful feature of Folio Views is that while creating a comprehensive base (appending all the fieldnotes and interviews of all the ethnographers), it breaks the text fields into 'idea-size' or paragraph-size segments: called a *Folio*. Breaking textual data into index cards is a method that survives the days when personal computers were not available and everything was done manually. The idea is to break the large texts into paragraphs, and then copy them on to index-cards. Folios are the functional equivalent of index cards. The size of a folio may be customized to fit researchers' need. For example, a fifty-page file, (written in WordPerfect) containing 150 paragraphs, may be converted into one folio or any number of folios. Furthermore, the database may be updated anytime. It can be appended, modified, and selected portions can be deleted and stored in a new or the same infobase. During the conversion process, the program automatically indexes (in alphabetical order) all the words, including articles, verbs, prepositions and numbers (separated by a space).

Before we started analysis, for the sake of consistency, all field notes and interviews by an ethnographer were chronologically organized. Some files were larger than others but most interviews were more than a hundred pages long. A comprehensive infobase was created by appending all bases together.

## Analysis

What do respondents mean by a given statement? Qualitative data presents the researchers with a multitude of messages and meta-messages. By its very nature qualitative data raises some hermeneutic issues. Unlike quantitative data, textual data always carry a message, even without any context. However, without a context the message becomes enigmatic. Thus, to decipher the 'meaning' of a statement requires interpretation.<sup>2</sup>

The ethnographic data consisted of texts full of respondents' words and ethnographers' descriptions. The usual method of making sense of textual data was by coding them uniformly. But the amorphous nature and multiple meanings in textual data greatly increased problems of coding. Two coders found it very difficult to interpret the texts with the same set of codes (if these could be created) and yet have the codes mean the same thing. Although we wanted to address several simple themes which could be framed as hypotheses, major problems arose in finding the relevant data. Such as we were looking at crack distribution pattern, but did not have any a priori idea about the pattern itself.

## Coding

Why code data at all? For analysis and interpretation coding is an essential step, especially when the respondents say one thing but mean something else. For example, if we know that 'blow-job' may lead to cash or payment in kind (crack) this is important for calculation of income or 'hustling' means income potential, so rather than coding them for income or earning we may directly look for these words. In that case rather than coding words, we could concentrate on coding the substantive research topics. The coding system should ideally be standard, but also acceptable to all the ethnographers and at the same time must be detailed enough to capture maximum possible information. Coding was intended to ease access to key

---

<sup>2</sup>Interpretation is used here in the sense Taylor uses for 'experiential meaning': "Interpretation, in the sense relevant to hermeneutics, is an attempt to make clear, to make sense of an object of study. This object must, therefore, be a text, or a text-analogue, which in some way is confused, incomplete, cloudy, seemingly contradictory — in one way or another unclear. The interpretation aims to bring to light an underlying coherence or sense" (Taylor 1987:33).

concepts for retrieval and to capture the words and phrases, including many variants of a single concept.

*Reading and identifying* substantive portions of data for analysis, from among twenty-five-thousand pages of electronic files or even printouts was not possible, not to mention identifying, marking, and coding them. The data needed to be coded so that easy retrieval was possible, and yet provide sufficient coverage of the many themes embedded in the infobase. Coding posed the usual methodological and logistic problems.

After consultation with ethnographers who conducted the interviews and the researchers who were involved in analysis, a code book was developed. The main purpose of the code book was to incorporate codes in the textual data for many different analytical themes without losing information contained in the text. The code book was developed through a continuous process of coding, followed by discussion, and refinement of coding categories. Over six months an acceptable system evolved into a 50-page coding manual. In its final form, the code book contained more than twenty-one categories, each section had nine subsections and each subsection with ten or more codes. The major categories were broad, e.g., Attitudes; Demographic background; Police activities; Economic behavior; and Family. The subsections in each category had many detail codes: Income from the drug sale, Income from legal employment. Income from other sources like stealing or hustling, etc.

The major problem with this (and any) coding system was interpretation of data. Coding textual data is virtually a process of interpretation — a process of translation. Like all translations, coding categories were insufficient to capture the full breath of meanings of specific words and phrases that many of us wanted to capture. Everyone involved in the process felt that the coding categories were either too broad or, if specific, did not capture the ‘essence’ of the information. Another problem arose: by fragmenting texts with fine codes, the main point of the sentence, paragraph, or a section of an interview was often missed. Everyone agreed that ‘the whole was bigger than the sum of its parts’.

### **Emic Analysis from the Subject's Viewpoint**

Before long we realized that some standard rules must be followed to be consistent when interpreting the data. The same information, the same answer to a particular question, may vary in meaning depending upon the person who said it and how, when, and in what context it was said. The question of standardized rules of interpretation is not only debatable but an unresolved issue.<sup>3</sup> Since meaning is predicated upon the subject, then questions like — “what do the ‘natives’ mean?” — remain problematic. This shift in the paradigm, from the outsider's point of view to the native's point of view, is a debatable issue.<sup>4</sup>

Examining the problem of drug distribution from the ‘emic’ view of the crack users/dealers meant understanding peoples' lives as they experienced them and not only describing their activities. To achieve this goal, we wanted to search and access emic concepts, words, ideas, phrases, and descriptions in the ‘native language’. ‘Emic’ terms require a different kind of data operation (coding/labeling, accessing, and collating) to trace these words, concepts, and descriptions; in other words, contextualizing them.

<sup>3</sup>However, as Taylor argues that there are three inseparable dimensions in the search for a meaning: “(1) Meaning is for a subject ... (2) Meaning is of something ... And (3) things only have meaning in a field, that is, in relation to the meanings of other things” (Taylor 1987:41).

<sup>4</sup>The debate about and analytic implications of emic and etic perspectives have expanded far beyond what Kenneth Pike first meant by these concepts: “The terms themselves were coined by the missionary linguist Kenneth Pike on analogy with the “emic” in phonemic and the “etic” in phonetic. In conformity with this analogy, Pike stressed “the structural results” obtained by phonemic analysis as opposed to the “nonstructural” results of phonetics” (Harris 1968:569).

## Etic Analysis from the Observer's Viewpoint

What does a researcher understand a statement or a set of statements to mean? If the agenda of social science is to understand human activities (including subjects' ideas) and make sense out of them, the interpretation remains constrained by our analytic paradigm, the etic analysis.

Although not conspicuously, the etic perspective dominates emic one in traditional anthropological analysis. When interpreting data, the analyst tries to translate information from one culture into another cultural frame, language, concept, diction, and norm. It is similar to an exercise in cross-cultural and intersubjective analytic scheme. In etic analysis, the analyst wants to trace patterns of cognition, behavior, and feeling among the subject population similar to or explainable by another set of concepts, his own. In this case researchers seek to retrieve some of the analytical words, concepts, ideas, and phrases that are traditionally used by social scientists. This requires a different kind of data operation.

Whether we wanted to prove or disprove some hypotheses, from either emic or etic viewpoint, information had to be located in the database. The main goal was to locate sections of the texts which were relevant to an analytic topic. Merged fields were essential to test a hypothetical relationship or develop a common trend across different subjects. Even simple analysis was not possible without sorting and retrieving data in some ordered fashion. For example, when we wanted to test an etic hypothesis: *All drug dealers are married*, or to develop an analytical theme: *Drug dealers/users were engaged in a dangerous business*, we had to look for critical information across the whole database. To search for data to support or discard the hypothesis, our analysis could go in two different directions. First, we could retrieve drug dealers' marital status to see if all of them were married or not. Since we did not ask about marriage and the sequence of questions varied from one ethnographer to the next, we retrieved texts where respondents talked about marriage, married life, family, husband, wife, spouse, children, son, daughter, or other related issues. Respondents could have discussed these issues in many different contexts. Manual searches for such data would have been gigantic tasks, with few guaranteed results.

For development of a thematic question, whether the drug business is dangerous or not, the search operation becomes more complicated. A search for texts where respondents talked about the drug business, drug dealing, trading, exchanges/barter, violence and the like, could be conducted, to find out if subjects viewed dealing as dangerous or not in those contexts. On the other hand, if we tried the emic perspective, different questions arose: Do the dealers themselves report their drug dealing activities to be dangerous? The question may be considered a false one imposed from outside. Usually, while searches would include references to the drug business, we needed to make a decision whether the information sought was from the point of view of the ethnographer or of the participants.

## Some Concluding Comments

In textual data, the search is not conducted for words as much as for information about a given concept/idea in order to develop an analytic theme; the search is for the meaning of the text that involves the choice of analytic perspectives, whether to choose emic or etic. For example, when a researcher wants to know about the income of drug dealers, what is it precisely that he is looking for? Is he trying to calculate earnings of a drug dealer from all possible sources? And if so, is he looking for it from the etic perspective or emic perspective? In a dealer's mind 'income' may mean something different from what the researcher may like to know about.

When we tried the etic perspective and searched for the words like 'salary', 'earning,' or 'income' no references could be found. This index-based search results almost mislead us to make a false conclusion that the drug dealers have no income or work for free, which was

contrary to our a priori knowledge that it was not true. This analytic failure, however, has nothing to do with the search for these words. But when we tried an emic perspective, we found that drug dealers are not *salaried employees*. First, this critical information is at the center of the social life of the dealers presenting a constraint in a word-based search in the text. The fact that drug users or dealers are not salaried employees (neither do they view their income and earnings as defined by an economist (etic) nor do they use the terms ‘income’ and ‘earning’ in their ordinary speech) cannot be deduced unless we shifted our analytic frame. If income, earning, and gains are not discussed in interviews, how then can we access this information? Secondly, in drug dealings, although street workers may get regular payments from their ‘bosses’, they may not call that ‘salary’, nor ‘income’ or ‘earning’. They may use street terms, e.g., how much they ‘hustled’ or ‘made’ on a particular day. Thirdly, a search for the index ‘salary’, revealed that ‘salary’ referred only to their pay from occasional legal jobs, but never for money received from selling drugs.

Since Folio Views captures all the words, we could utilize both perspectives searching for words and meanings in the text. Beyond the emic-etic dilemma, another significant aspect of data analysis is the social life of the drug dealers. We also found several female dealers in the crack business report earnings from prostitution, a whole series of street-words like ‘freaking-out’, ‘blow-job’, ‘half-and-half’, and ‘straight and kinky sex’ were more revealing than the word ‘prostitution’.

Analytically, the relationship between ‘income’ and ‘earning’ on the one side and ‘rapping’, ‘cooking’, and ‘sex for crack’ on the other is not a semantic relationship — these words are not synonymous. The relationship among the words lies in the social realm and implicit paradigms of the people. Decontextualizing the words and quotes from the social context may lead to spurious analysis. The real meanings of words are built upon the social structure and normative framework of the target population. Whether a researcher is conducting a preconceived search (to test a hypothesis or develop a theme) or searching for a heuristic meaning (patterns, relationships, and synthesis) to emerge, his analysis is dependent on the understanding of the target population. The analysis of the textual data remains grounded in the *social* rather than in the *semantic* sphere. If the social parameters of the structure are unknown, neither emic nor etic meanings may be established. Ethnographic text in emic language and diction is a cultural product of the people. This assertion is based on the following facts. First, at the cognitive level, textual discourse represents the life of the people — those who are speaking and constructing it. The discourse construction by drug dealers often defies the standard rules of grammar and standard word meaning. Drug users/abusers/dealers talk in a metaphoric language and coin words that do not mean anything to people who do not belong to their culture. Secondly, the discourse also embodies the wishes, desires, and real conditions of the people. In the voice of the respondents the ‘native point of view’ emerges as the only way of achieving authenticity (Geertz 1973). The combination of these two analytic trajectories (what and how they are saying) offers a reflexive total view that researchers are only beginning to discern. The strength of ethnographic data is in its integrity, simplicity, and authenticity. The data urged us to be faithful to both dimensions; rather than imposing arbitrary meanings from outside, the important thing about lives of drug dealers was not what it seemed like, but how it worked.

## Acknowledgments

This research was supported in part by Natural History of Crack Distribution (1 RO1 DA05126-02) and Behavioral Sciences Training in Drug Abuse Research Program (5 T32 DA07233-07). Additional support was provided by National Development and Research Institutes, Inc. Points of view and opinions in this paper do not necessarily represent the official positions of the United States Government, Medical and Health Research Association of New York City, Inc. or National Development and Research Institutes, Inc. We thank with appreciation Dr. Ansley Hamid and the three anonymous reviewers for their comments on an earlier version of this paper.

## References

- Adler, Patricia. *Wheeling and Dealing: An Ethnography of an Upper-Level Drug Dealing and Smuggling Community*. New York: Columbia University Press; 1985.
- Geertz, Clifford. *Interpretation of Cultures*. New York: Basic Books; 1973.
- Harris, Marvin. *The Rise of Anthropological Theory*. New York: Thomas Y. Cromwell Company; 1968.
- Hymes, Del, editor. *Reinventing Anthropology*. New York: Vintage Books; 1974.
- Kotarba, Joseph A. Ethnography and AIDS: Returning to the Streets. *Journal of Contemporary Ethnography* 1990;19(3):259–270.
- Marcus, George E.; Dick, Cushman. Ethnographies as Texts. *Annual Review of Anthropology* 1982;11:25–69.
- Taylor, Charles. Interpretation and the Sciences of Man. In: Rabinow, Paul; William M, Sullivan, editors. *Interpretive Social Science: A Second Look*. Berkeley: University of California Press; 1987.