Qualitative Data, Quantitative Analysis
H. Russell Bernard
*Field Methods* 1996 8: 9
DOI: 10.1177/1525822X960080010401

The online version of this article can be found at:
http://fmx.sagepub.com/content/8/1/9

Published by:
http://www.sagepublications.com

Additional services and information for *Field Methods* can be found at:

Email Alerts: http://fmx.sagepub.com/cgi/alerts

Subscriptions: http://fmx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://fmx.sagepub.com/content/8/1/9.refs.html

>> Version of Record - Feb 1, 1996

What is This?
self because I kind of 'did it' for no reason. But when he does stick around with (the baby) for me, it worked. But it's really hard when you're pregnant or have a baby to even get someone to be with you. I don't want to do this by myself anymore, you know. So I try not to.

Discussion

This essay shows the need to examine the motives behind street youths trading sex for things other than money, drugs, and alcohol. In particular, we need to know much more about the prevalence of trading sex for physical security, job security, and child care. Are such trades exclusive to the group I studied, or are all street adolescents forced by similar circumstances into similar behavior?

If trades of sex for physical security and child care are, in fact, common among street youth, then programs to reduce this behavior should provide adolescents with free child care and accessible shelters where they can escape physical threats. Increasing the policing of areas where street adolescents are physically threatened would also reduce their need to trade sex for security. Additional research on these activities would benefit from combining survey data with an ethnographic component, thus giving a more complete picture of risky behaviors.

References


CAM • February 1996

Qualitative Data, Quantitative Analysis

H. Russell Bernard
University of Florida
ufruss@nervm.nerdc.ufl.edu

The Human Relations Area Files (HRAF), for example, consists of about one million pages of text on about 550 societies around the world. The data on a 60–culture sample from that database are available on CD-ROM and HRAF is converting the entire million-page corpus of text to machine-readable form. The Center for Electronic Texts in the Humanities at Rutgers University is bringing together machine-readable corpora (Shakespeare’s works, ancient Greek and Latin literature). Legal scholars can access some court opinions on line. The list goes on and on. Conversions of text corpora to on-line databases are proceeding at a breathtaking pace.

Qual/Quant and Texts

One of the things I like best about texts is that they are as valuable to positivists as they are to interpretivists. Positivists can tag text and can study regularities across the tags. This is what content analysis (including cross-cultural hypothesis testing) is about. Interpretivists can study meaning or look for the narrative flourishes that authors use in the (sometimes successful, sometimes unsuccessful) attempt to make texts convincing.

Scholars of social change have lots of longitudinal quantitative data available (the Gallup poll for the last 50 years, baseball statistics for over 100 years, to name a couple of well-studied data sets), but longitudinal text data are produced naturally all the time. For a window on American popular culture, take a look at the themes dealt with in country music and in Superman comics over the years.

Or look at sitcoms and product ads from the 1950s and from the 1990s. Notice the differences in, say, how women are portrayed or what people think is funny in different eras. In the 1950s, Lucille Ball created a furor when she became pregnant and continued making episodes of the I Love Lucy show. Now think about almost any episode of Seinfeld. Or scan some of the recent episodes of popular soap operas and compare them to

Continued on page 10

Page 9
those of 30 years ago. Today’s sitcoms and soaps contain much more sexual innuendo.

How much more? To measure that, you could code a representative sample of exemplars (sitcoms, soaps) from the 1950s and another representative sample from the 1990s, and compare the codes (content analysis again). Interpretiveists might be more interested in understanding the meaning across time of concepts like “flirtation,” “deceit,” “betrayal,” “sensuality,” and “love,” or the narrative mechanisms by which any of these concepts is displayed or responded to by various characters.

Suppose you ask a hundred mothers to describe their last pregnancy, or a hundred labor migrants to describe their last (or most dangerous, or most memorable) illegal crossing of the border, or a hundred hunters to describe their last (or most difficult, or most thrilling) kill. In the same way that a hundred episodes of soap operas will contain patterns about culture that are of interest, so will a hundred texts about pregnancies and hunts and border crossings.

The Coding Problem

The difficulty, of course, is in coding texts and in finding patterns. Coding turns qualitative data (texts) into quantitative data (codes), and those codes can be just as arbitrary as the codes we make up in the construction of questionnaires.

When I was in high school, a physics teacher put a bottle of Coke on his desk and challenged our class to come up with interesting ways to describe that bottle. For weeks it sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk. For weeks it sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk. For weeks it sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk. For weeks it sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk. For weeks it sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk.

Cell a is the qualitative analysis of qualitative data. Interpretive studies of texts are of this kind. Studies of the cell d variety, by contrast, involve the statistical analysis of questionnaire data, as well as more mathematical kinds of analysis.

Cell b is the qualitative analysis of quantitative data. It’s the search for, and the presentation of, meaning in the results of quantitative data processing. It’s what quantitative analysts do after they get through doing the work in cell d. Without the work in cell b, cell d studies are puerile.

Which leaves cell c, the quantitative analysis of qualitative data. This involves turning the data from words or images into numbers. Scholars in communications, for example, might tag a set of television ads from Mexico and the U.S. to test whether consumers are portrayed as older in one country than in the other. Political scientists might code the rhetoric of a presidential debate to look for patterns and predictors. Archeologists might code a set of artifacts to produce emergent categories or styles, or to test whether some intrusive artifacts can be traced to a source. Cultural anthropologists might test hypotheses across cultures by coding data from the million-pages of ethnography in the Human Relations Area Files and then doing a statistical analysis on the set of codes.

Strictly speaking there is no such thing as a quantitative analysis of qualitative data. The qualitative data (artifacts, speeches, ethnographies, TV ads) have to be turned first into a matrix, where the rows are units of analysis (artifacts, speeches, cultures, TV ads), the columns are variables, and the cells are values for each unit of analysis on each variable.

On the other hand, the idea of a qualitative analysis of qualitative data is not so clear-cut, either. It’s tempting to think that qualitative analysis of text (analysis of text without any recourse to coding and counting) keeps you somehow “close to the data.” When you do a qualitative analysis of a text, you interpret it. You focus on and name themes and tell the story, as you see it, of how the themes got into the text in the first place, perhaps by telling your audience something about the speaker whose text you’re analyzing. You talk about how the themes are related to one another. You may deconstruct the text, look for hidden subtexts, and in general try to let your audience know the deeper meaning or the multiple meanings of the text.

In any event, you have to talk about the text, which means you have to produce labels for themes and labels for articulations between themes. All this gets you away from the data, as surely as numerical coding does. Qualitative analysis involves reducing people (as observed directly or through their texts) to numbers, while qualitative analysis involves reducing people to words.

I don’t want to belabor this, and I certainly don’t want to judge whether one reduction is better or worse than the other. Scholars today have at their disposal a tremendous set of tools for collecting, parsing, deconstructing, analyzing, and understanding the meaning of data about human thought and human behavior. Different methods for doing these things leads us to different answers, insights, conclusions and, in the case of policy issues, actions. Those actions have consequences, irrespective of whether our input comes from the analysis of numbers or of words.

Note

1. This was written while I was at the University of Cologne (July 1994–July 1995). I thank the Alexander von Humboldt...
References


Continued from page 12

you want to know. In these cases pauses have the shape of words between spaces. If special characters are treated by the computer program as if they were alphabetic characters, text analysis procedures will work.

4. Functional Notation

#(time) very long pauses, giving time in seconds, e.g. #(25) for a 25-second pause. Could be used for fractions of a second as well.

RSH(text) pitch register shift up (register shift high; ( start, ) end)

RSL(text) pitch register shift down (register shift low)

ACC(text) accelerated tempo

ACC/RSH(text) accelerated tempo with register shift high over text

Thus combinations of functions or nested functions are possible.

DEC(text) decreased tempo

FOR(text) fortis enunciation (spoken sharply)

LEN(text) lenis enunciation (spoken softly)

LOU(text) loud enunciation (spoken loudly; can be nested for increasing loudness)

SOF(text) soft enunciation (spoken quietly, can be nested for decreasing loudness)

WIS(text) for a whisper

ALT(word or text) alternative hearing of a word or phrase

POS (word or text) possible hearing

Alternative: The following convention has the advantage that it is more intuitive than the functional notation. Unfortunately this “analog” method works only with loudness and with none of the other functions.

No louder (larger type, or increasing type size)

No less loud (smaller type, or decreasing type size)

The functions are capitalized to set them apart from transcribed text. The parenthesized text determines the range of the modifier function: the left parenthesis marks where the modification starts, the right marks its end. Thus WIS(no) implies a whispered no. In any program that treats parentheses as non-letters these function words are sorted with all other words. Problems arise if an ethnographer wants, for example, a count and display of all loud passages.

5. Breath

.hh in-breath

.hh out-breath (the length of the rows of h indicates the length of the out-breath, each h about 1/4 second)

The assumption is that the h’s are flanked by the space characters and thus do not not interfere with word sorts or frequency counts.

6. Over-Speech

For conversational overlap (simultaneous speech of two or more speakers), use two or more lines with vertical alignment at places where overlaps between speakers start and end. The analogy here is musical notation, where each instrument has its line of notes. Text processing programs generally ignore the extra spaces. There are few programs available that allow separation of the output of different speakers—for example (and most simply), ethnographer and consultant.

In many situations overspeech may be difficult to unscramble. Stereo recording makes unscrambling and multiple speakers’ identification easier to manage.

7. Unintelligible Words

XXX use three capital X’s for unintelligible words and groups of three X’s to indicate the approximate length of an entire unintelligible utterance. A count of the XXX forms should indicate the approximate number of unintelligible words in a text.

8. Commentary

[] May contain any kind of commentary by interviewer, transcriber, analyst, or ethnographer. Use these brackets to insert comments and descriptions of the conversational setting, for example.

Conclusion

The marking of discourse features is important for ethnography. However the notation calls for improvement shaped by experience. I hope that readers will react to my suggestions so that we can jointly develop a better, simpler, more adequate system.

Note

1. I think that a quarter second is an interval any human being can estimate reasonably accurately. Silverman’s (1993:118) 1/10 second intervals become practical only with a stop watch.

References


