Using Methods in the Field

A Practical Introduction and Casebook

Victor C. de Munck
Elisa J. Sobo

Editors

ALTAMIRA PRESS

A Division of Sage Publications, Inc.
Walnut Creek ■ London ■ New Delhi

For information address:

AltaMira Press
A Division of Sage Publications, Inc.
1630 North Main Street, Suite 367
Walnut Creek, CA 94596
http://www.altamirapress.com

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash 1
New Delhi 110 048 India

PRINTED IN THE UNITED STATES OF AMERICA

Library of Congress Cataloging-in-Publication Data

Using methods in the field: a practical introduction and casebook / edited by Victor C. de Munck, Elisa J. Sobo.
p. cm.
Includes bibliographical references and index.
ISBN 0–7619–8912–9 (cloth)
ISBN 0–7619–8913–7 (pbk).
1. Ethnology—Methodology. 2. Social sciences—Methodology. I. Munck, Victor C. de. II. Sobo, Elisa Janine, 1963–
GN345 .138 1998
305.8'001–ddc21 98–9073
CIP

98 99 00 01 02 9 8 7 6 5 4 3 2 1

Editorial Management by Jennifer R. Collier
Production Services by Carole M. Bernard, ECS
Cover Design by Eric Akeson

GERY RYAN ■
THOMAS WEISNER

Three

Content Analysis of Words in Brief Descriptions: How Fathers and Mothers Describe Their Children

Introduction

Our objective in this study is to understand parental perceptions and attitudes of Americans toward their adolescent children. From our background readings in anthropology, we know that parental impressions of adolescents and children vary across cultures. When Super and Harkness (1986) asked Kipsigi parents in western Kenya to describe boys, descriptions included the terms "warrior" and "fierce." When Raghavan (1993) asked South Asian parents living in the United States about their daughters, descriptions included "hospitable" and "responsible." Such phrases or words would strike most American parents as unusual or odd. American parents are more likely to use such terms as "athletic," "independent," "argumentative," and "well-rounded"—terms that would seem odd to most Kipsigis or South Asians.

The data for this article come from a follow-up survey in the Family Lifestyles Project—a 20-year longitudinal study of conventional and nonconventional families in the United States.[1] As part of the survey, we asked parents: "What is your teenager like now? Does she or he have any special qualities or abilities?" This was a relatively easy, comfortable task for most people we interviewed. Parents wrote their answers in short descriptive phrases. We focus now on these data.

We could have asked many questions about these descriptions. For our research purposes, however, two were particularly important: "How did parents describe their children?" and "Did mothers and fathers describe their children differently?"

In 82 of the 200 families interviewed, a male and female parent **independently** described their child. Nearly all the descriptions came from the biological parents. In three cases, the male parent was a step-father, and in one case the female parent

was a step-mother. Because we are interested in how parents raise their children, we treat biological and step-parents as equal in our analysis. We thus have two descriptions for each of 82 children, one from the mother and the other from the father, for a total of 164 descriptions.

Each of the 82 children are different (some are more artistic, social, academic, or temperamental than others), but we can make comparisons across children because: (1) We were systematic in how we asked parents to describe their experiences (we always asked the exact same question each time); (2) Each pair of parents described the same child; and (3) We have the same number of descriptions (82) from mothers and from fathers. Of course, the child is not "the same" to each parent. Children respond and identify differently to each parent and parents to each child. So, when parents are asked to describe their child, they are not reacting to exactly the same stimuli, but to a comparable family situation that has different meanings to each family member.

Selecting an Analysis Strategy

Here is a classic example of anthropologists collecting good, rich, ethnographic data from their informants, and deciding how to analyze it afterward. Usually, you want to know how you are going to analyze your data before you collect it. Choosing an analysis strategy is analogous to planning a road trip. You need to know first where you are starting from (what kind of data you have) and then you need to know where you want to go (what research goals you have). We had textual data that came from **open-ended questions**. This meant that respondents were not confined to a fixed set of predetermined answers. They simply told us what they thought in their own words. As analysts, we had two goals. We wanted to *describe* the kinds of answers that parents gave and we wanted to *compare* fathers' and mothers' answers.

The analysis strategy that best links open-ended textual data with the research goals requiring a systematic description and comparison is loosely referred to as *content analysis*. There are several flavors of content analysis (see Bernard and Ryan [1998] for a broad overview). Classic content analysis is based on judgments the researcher makes. In this case, the researcher identifies a fixed set of themes or categories, then reads each text and marks down how often the text refers to each theme. Sometimes, they mark down whether the theme is mentioned anywhere in the text; at other times, they count the number of words or paragraphs that pertain to a theme. During World War II, when the approach was first developed to analyze German propaganda, coders actually measured the number of column inches in newspapers dedicated to particular themes.

Because of the increased use of computers due to their improvements and lower costs, a new type of content analysis has appeared. In this approach, the researcher

doesn't pick a fixed set of themes or categories, but analyzes the words used by the informants. How much can we learn from a simple word analysis of qualitative data? Jehn and Doucet (1996, 1997) used word lists derived from informants' descriptions of recent confrontations to learn to better understand inter- and intracultural conflicts. Schnegg and Bernard (1996) relied on computer-generated word counts to identify central themes that graduate students in anthropology used when describing their own fields of study. Jang and Barnett (1994) were able to distinguish between American and Japanese companies based on the words used in each company's published annual reports. (For additional examples of content analysis, see Krippendorff [1980], Weber [1990], and Roberts [1997].)

We rely on standard word-processing programs and other readily available software. No special formatting or coding is needed. The methods we describe are useful for discovering patterns in any body of text, whether field notes or responses to open-ended questions, and are particularly helpful when used along with ethnographic data and other sources of information. Word analysis can tell us about a word's salience. By "salience" we mean the importance of a word as measured by the number times it is used informants. We can also use word analysis to examine word patterning, where "patterning" refers to how often pairs of words are found close together (for example, within the same sentence, paragraph, or interview). Word analysis, however, cannot produce a holistic interpretation of cultural data because it takes the words out of their original context.

In our data, we asked two questions: (1) What do the words parents use in their descriptions tell us about the goals they have for their children?; and (2) What do differences and similarities in word use tell us about the differences and similarities in parents' perceptions of their children?

Organizing the Data

We transcribed the parents' verbatim answers into a word processor text file (in this case, WordPerfect 6.0). For each answer, we typed in the family identification number, the type of family, the sex of the child being described, the sex of the parent who gave the description, and the complete description. Each description was followed by a single hard return. Figure 3.1 shows the first three descriptions in our master file (MASTER.WP).

To facilitate analysis, we separated each unique phrase/descriptor by a period and a space. The period/space combination has two advantages: (1) A period indicates the end of a sentence, and we can then use the word processor or style checker to count the number of sentences in a document (Harris 1996); (2) We can use the period as a delimiter for importing the text data into a spreadsheet or a database (like Excel or Quattro Pro).

ID009. F1030. Boy. Fthr. Loving. Obedient. Maintains own identity. Likes being home. Independent. Anxious to go to California to school.
ID016. F1130. Boy. Fthr. Smart. Energetic. Arrogant. Dependent. Slick. Passive. Lack of imagination. Attraction to inner-city lifestyle.
ID124. F1130. Girl. Mthr. Great kid. Willing to communicate with parents. Listens. Motivated in school. Helpful around the house. Healthy. Active. Lots of friends. Tends to play it safe.

Figure 3.1. Examples of master file of parents' descriptions of their children.

Once we had our master file of descriptions, we sorted the descriptions by parent's sex. Since we consistently made parent's sex the fourth word of the paragraph, we can do this with our word processor. Select all your text and tell the word processor to use the fourth word to sort the highlighted paragraphs.[2] (Before sorting, back-up your file!)

We then copied mothers' and fathers' responses to separate files (MOTHER.WP & FATHER.WP). At this point, we were only interested in the descriptors, so we stripped out the extraneous information in each file. This is easily semiautomated with a **macro** that goes to the beginning of each paragraph and deletes the first four words (ID, family type, child's and parent's sex). Our two stripped files contained only the verbatim descriptions provided by mothers and fathers.

Simple Tricks with a Word Processor

We used WordPerfect's document information function to calculate some general statistics.[3] Document information is located under File on the top menu. Among other things, it calculates the number of characters, words, and sentences, plus the average word length, the average number of words per sentence, and the maximum words per sentence. Table 3.1 compares these statistics for mothers' and fathers' responses.

These statistics tell us that:

1. Mothers use more words to describe their children than do fathers. Of all the words used to describe the 82 children, 56% come from mothers and 44% come from fathers.
2. On average, mothers used 28% more sentences than did men. (Mothers used 528/82 = 6.4 phrases to describe their children, while men used 411/82 = 5.0 phrases. Mothers and fathers use the same number of words per phrases, but mothers said more things about their children.)
3. Mothers and fathers use roughly the same size words, about 5.7 characters each.

TABLE 3.1
Text Statistics Generated from WordPerfect 6.0

	Mothers	Fathers	Total
Characters	9,748	7,625	17,373
Word Count	1,692	1,346	3,038
Sentence Count	528	411	939
Average Word Length	5.76	5.66	5.72
Average Words per Sentence	3.20	3.27	3.24
Maximum Words per Sentence	14	17	17

Fathers and mothers are more similar in this sample than they are different. Mothers use more words, but not very much more, and on other measures, fathers and mothers are about equal. Clearly, parents used the same "standard social science questionnaire schema" to answer our questions—writing a series of terse phrases and words for a minute or so.

Learning from Unique Word Lists

We next examine whether mothers and fathers use *different* words to describe their children. WORDS 2.0 (E. Johnson 1995) is a useful program that counts the number of running words in a text, identifies the number of **unique words forms**, and lists the number of occurrences of each unique form.[4] (See Bernard [1995] for a review of WORDS 2.0.) Other programs, such as CATPAC, and the latest version of ANTHROPAC (Borgatti 1992) also count the frequency of unique words. (See Doerfel and Barnett [1996] for a review of CATPAC.)

To get the files ready for WORDS 2.0, we first saved our WordPerfect files (MOTHER.WP and FATHER.WP) in ASCII format (calling them MOTHER.ASC and FATHER.ASC so as not to overwrite the original files). When we analyzed each file, we used WORDS 2.0's "**common word list**" to exclude 125 of the most-used English terms. Figure 3.2 shows a portion of the two outputs. Each output tells us how many words each file contained originally,[5] how many unique words were found (including unique common words), and how many words were removed when we eliminated the common ones. WORDS 2.0 outputs the list of unique words with their respective frequency of occurrence. We indicate the rank order of each word under the # sign. (You can do this in your word processor by turning on the line numbering option.)[6]

Figure 3.2 shows that the MOTHER file contained a total of 1,721 words in 734 unique word forms. It contained 542 instances of the 125 common words that were eliminated from further consideration. In the end, there were 666 unique words in the file, and mothers mentioned the words "good," "friends," "loving,"

Mothers' Descriptions			Fathers' Descriptions		
File MOTHER.ASC:			File FATHER.ASC:		
Total number of running words in file: 1,721			Total number of running words in file: 1,355		
Number of unique word forms in file: 734			Number of unique word forms in file: 607		
The following counts exclude 542			The following counts exclude 419		
occurrences of 125 common word forms.			occurrences of 125 common word forms.		
Rank	Frequency	Word	Rank	Frequency	Word
1	22	good	1	23	good
2	12	friends	2	16	school
3	11	loving	3	11	hard
4	11	out	4	9	intelligent
5	11	people	5	8	bright
6	10	doesn't	6	8	independent
7	10	hard	7	8	out
8	10	school	8	8	well
9	9	responsible	9	7	doesn't
10	9	sense	10	7	lack
11	8	caring	11	7	loving
12	8	intelligent	12	7	people
13	8	lacks	13	7	sensitive
14	8	sensitive	14	7	sports
15	7	bright	15	7	student
16	7	honest	16	6	caring
17	7	others	17	6	does
18	7	self	18	6	life
19	7	time	19	6	others
20	7	well	20	6	work
21	7	work	21	5	ability
22	6	creative	22	5	enjoys
23	6	does	23	5	great
24	6	great	24	5	lacks
25	6	mature	25	5	likes
26	6	sports	26	5	mature
27	5	academically	27	5	own
28	5	artistic	28	5	sense
29	5	cares	29	5	social
30	5	concerned	30	5	wants
31	5	goals	...	...	...
32	5	going	548	1	zero
33	5	humor			
34	5	independent			
35	5	other			
36	5	social			
37	5	times			
...	...	...			
666	1	zest			

Figure 3.2. Counts of words used more than five times by mothers and fathers.

"out," and "people" at least 11 times. The last word on the mothers' list, "zest," was mentioned only once.

We can think of unique word lists as *concentrated* data or, as Tesch (1990: 138–139) called them, distillations. **Concentration rates** are also called **type-token rates**. Concentration rates refer to the average rate at which words in a text are used multiple times. The rates are calculated with the formula: 1– (total unique words/total words). We can produce three different measures of concentration and can compare any of these measures across the MOTHER and FATHER files.

The first concentration rate uses *all* the unique words (regardless if they are common words or not) in the calculation. With 734 unique words in a corpus of 1,721 words, mothers have a concentration rate of 57% (1–734/1,721). Fathers have a concentration rate of 55% (1–607/1,355). The second concentration rate doesn't include the common words in its calculation. For mothers, we calculate the new concentration rate based on the 666 unique substantive words (eliminating all occurrences of words in the common-word file). Thus, the rate for mothers is 1–666/1,721 = 61%. For fathers the rate is 1–548/1,355 = 60%. We also can calculate a third concentration rate based on words that occur more than once. In the MOTHER file, just 207 of the 666 substantive words occur more than once. This produces a concentration rate of 1–207/1,721 =88 %, identical to the rate (1–159/1,355 = 88%) for fathers.

We lose a lot of information when we examine unique words. We don't know the context in which the words occurred, nor whether informants used words negatively or positively. Nor do we know how the words related to each other. But distillations like these introduce very little **investigator bias** (we *do* have to choose what words to leave out of the analysis), and they can help us identify constructs used by parents to describe their children.

The word lists suggest things about parents' values and goals for their children, and the lists can be compared across fathers and mothers. For example, from Table 3.1 we don't know if fathers have less to say about their children or they just have less to say about all topics. From Figure 3.2, however, we see that men's vocabulary for describing children is as rich as women's vocabulary. [The ratio of unique words to total words is roughly equivalent for men (607/1,355 = .45) and for women (734/1,721 = .43).]

Figure 3.2 lets us make crude comparisons between men's and women's use of different words. (The measures are crude because they represent **rank order data** and don't take into consideration the total number of words used by each group.) Both mothers and fathers use the word "good" a lot more than any other word. Women, for example use that word almost twice as much as "friends," their second most popular word. Antonyms of good aren't prevalent among the word list, indicating that people might have a tendency to be optimistic in describing their children, have a response bias on questionnaires to use positive words, and are

accessing a cultural model for describing one's child that emphasizes positive, growing cultural careers.

Figure 3.2 also suggests that men and women focus on different characteristics of their children. A comparison of the most frequently used words shows that "friends," "loving," "people," and "responsible" are ranked higher for women than they are for men. In contrast, "school," "hard," "intelligent," "bright," and "independent" are ranked higher for men than for women. This suggests that mothers, on first mention, express concern over interpersonal issues, while men appear to give priority to achievement-oriented and individualistic issues.

Figure 3.2 is very informative but is somewhat deceptive. For both men and women, the most frequently mentioned word was "good." Women mentioned it 22 times, and men mentioned it 23 times. This similarity in frequencies at first *seems* to suggest that men and women and women are equally likely to use that word to describe their children. But here lies the deception. Word frequencies don't consider the total number of words mentioned by men and women. (Remember, women used more words than did men!) To eliminate this deception, we can **standardize** the word frequencies according to what we expect to find if men and women used the same number of words. Standardize means to make two measures equivalent so they can be compared. Here we want to compare the use-rates of words across genders. Table 3.2 shows the results of such a process.

To create this table, we put mothers' and fathers' responses in a single ASCII file and counted the words again. We then selected the 131 words that informants mentioned at least four times. We picked four because it gave us a reasonable number of words to work with. (If we had selected three as our cut-off point, we would have had to work with over 200 words—something that seemed too daunting.) We put the these 131 words in the first column of a spreadsheet and their frequency counts in the second column. In the third and fourth columns, we put the number of times each word was mentioned by women and by men respectively. Next we calculated the *expected* word frequencies for men if men used the same number of words as women. Since women on average used 1.27 (1721/1,355) times more words as did men, we multiply the observed mens' frequency by 1.27. We put the result in the fifth column.

To compare mothers' and fathers' word use, we subtracted the expected mens' frequencies in column five from the observed women's frequencies in column three. We put the results in column six. This gave us a more accurate difference in word use between the files. Negative numbers in column six mean that the word was more likely to be used by fathers than used by mothers. Positive numbers mean that the word was more likely to be used by mothers. Numbers close to zero mean that there wasn't that much difference between the men's and women's descriptions. Finally, we sorted the rows in the spreadsheet by the values in column six. In Table 3.2, we show some selected results of this comparison: Words whose frequencies varied a lot between men and women as well as some examples of words whose frequencies varied little.

TABLE 3.2
Word Frequencies Sorted by Standardized Frequency Difference in Gender

TERM	Both	Mother	Father	Expected Father	Standardized Differences
school	26	10	16	20.3	−10.3
good	45	22	23	29.2	−7.2
lack	9	2	7	8.9	−6.9
student	9	2	7	8.9	−6.9
enjoys	6	1	5	6.4	−5.4
independent	13	5	8	10.2	−5.2
extremely	4	0	4	5.1	−5.1
like	4	0	4	5.1	−5.1
ability	7	2	5	6.4	−4.4
own	7	2	5	6.4	−4.4
wants	7	2	5	6.4	−4.4
high	5	1	4	5.1	−4.1
interested	5	1	4	5.1	−4.1
great	11	6	5	6.4	−0.4
mature	11	6	5	6.4	−0.4
humor	9	5	4	5.1	−0.1
times	9	5	4	5.1	−0.1
attitude	7	4	3	3.8	0.2
caring	14	8	6	7.6	0.4
adult	4	4	0	0.0	4.0
average	4	4	0	0.0	4.0
difficulty	4	4	0	0.0	4.0
goes	4	4	0	0.0	4.0
kid	4	4	0	0.0	4.0
lots	4	4	0	0.0	4.0
respect	4	4	0	0.0	4.0
talented	4	4	0	0.0	4.0
uses	4	4	0	0.0	4.0
honest	9	7	2	2.5	4.5
time	9	7	2	2.5	4.5
creative	6	6	0	0.0	6.0
friends	16	12	4	5.1	6.9

Now look at the word "good" in Table 3.2 (the second line). The table shows that if men and women used the same number of words in their descriptions, we would expect "good" to be used more often by men (29 times) than by women (22 times). This finding is different from those in Figure 3.2 that suggest that "good" is used equally by both men and women, By standardizing the data, we discover that there is a gender distinction that we would not have discovered otherwise.

Our standardizations in Table 3.2 tell us that fathers use the words "school," "good," "lack," "student," "enjoys," "independent," "extremely," "like," "ability," "own," "wants," "high," and "interested" more than do mothers. On the other hand, mothers use the words "friends," "creative," "time," "honest," "uses," "talented," "respect," "lots," "kid," "goes," "difficulty," "average," and "adult" more than do fathers. Men and women, however, are equally likely to use the words "great," "mature," "humor," "times," "attitude," and "caring."

Notice the differences between the standardized measures in Table 3.2 and the rank orders (and frequencies) shown in Figure 3.2. The rank-order data tell us about the relative priority of words within each gender, while the standardized data let us compare use rates across genders. For instance, the word "good" was the most-used word for both women and men. When we compare across genders, we find that men tend to use the word more often. In contrast, men and women are equally likely to use the word "caring" in their descriptions, but the simple rank order for the word is higher for women than it is for men. The next question we might want to ask is why do such gender differences and similarities occur?

Our findings are similar to other research on gender differences. On many measures, men and women, boys and girls show substantial overlap in behavioral tendencies. Although mean or modal differences often are relatively small, specific measures (in our case, emphasis on different concerns in describing teens) are quite constant and are found cross-culturally (Best et al. 1994).

The word-counting techniques described here don't require complex and expensive text analysis programs. These simple methods help researchers concentrate often confusing data into a more manageable form and are relatively bias free. The techniques can be used for exploring central themes and for systematically comparing within and across groups.

Of course, these are just the first univariate, exploratory steps in a more detailed qualitative analysis. We still want to examine the context in which these words occur and how key words are related to each other. For example: How does the sex of the teen as well as the parent influence word use? We also want to explore some of the hypotheses that we have formed in this simple first step. Treating words as units of analysis offers researchers a simple way of exploring text and confirming hypotheses.

QUESTIONS

1. How were the sample of texts selected? (Where did they come from? How many texts were analyzed? What percent were from mothers? What texts were not analyzed?)
2. What are the advantages and disadvantages of analyzing words instead of the whole text?
3. What do we learn about the difference between men and women when we look at the statistics associated with the total number of words, sentences, average word length, and average words per sentence?
4. How did the authors standardize men's in Table 3.2? Why did they do this?
5. How are men and women different in the words they use to characterize their children? How are they similar?

NOTES

1. The Family Lifestyles Project (FLS) started in 1974 when investigators contacted 200 mothers during their third trimester of pregnancy. Mothers were involved in conventional and nonconventional living arrangements. Nonconventional arrangements included single mothers, social contract couples (not legally married), and mothers in communes or group living situations. Members of the research team have followed the mothers, their mates, and their child ever since. Over the years, **attrition** has been remarkably low. In 1992–1994, the FLS researchers conducted a follow-up study of the adolescent children and reached 100% of the mothers, 98% of the teenagers, and 48% of the fathers or other mates. The central question of the adolescent follow-up was: How did these "children of the children of the '60s" turn out? To find out, investigators mailed parents a written questionnaire in which they asked parents about their child's performance in school, personal relationships, political attitudes, gender identity, drug use, and other characteristics. (See Eiduson and Weisner [1978], Weisner [1986], Weisner and Garnier [1992], and Weisner et al. [1994] for reviews and key findings from the project.)

2. To do this in WordPerfect for Windows 6.1: Select Tools/Sort from the menu. When the menu for sorting appears, tell WordPerfect to sort by paragraph. (Note: WordPerfect assumes that paragraphs are separated by *two* hard returns.) Make sure that the appropriate settings are marked as follows: Type = Alpha, Sort Order = Ascending, Line = 1, Field = 1, and Word = 4. After making the changes, select OK. WordPerfect will put all the fathers' responses on top of the file and all the mothers' responses on the bottom. To do this in Word 6.0: Select Table/Sort Text. Select Options. In the "Separates fields at": dialogue box, select "Other" and fill in the box with a single period. Select OK. In the "Sort by" dialogue box, select "Field 4." Select OK.

3. Similar statistics can be obtained in Microsoft's *Word*. *Word*, however, doesn't automatically count the number of sentences in a document. To do so, you need to build a macro, as follows:

```
Sub MAIN StartOfDocument Count = 0 While SentRight(1, 1) <> 0 If
Right$(Selection$(), 1) <>Chr$(13) Then count = count +1 Wend MsgBox "Number
of sentences in document:" + Str$(count) End Sub
```

4. WORDS 2.0 was created by Eric Johnson and is distributed by TEXT Technology, 114 Beadle Hall, Dakota State University, Madison, SD 57042–1799. Email: langners@columbia.dsu.edu. For information on other programs that Johnson has created, check out the website http://www.dsu.edu/~johnsone/ericpgms.html.

5. The total number of words identified by WordPerfect 6.0 for the MOTHERS.WP file (1,692) differs from the total number of words identified by WORDS 2.0 (1,721). For the same file, Word 6.0 counts 1,731 words. Discrepancies occur because each program has a slightly different definition of what counts as a word. In our case, single hyphens (-) are the leading culprits. WordPerfect 6.0 counts the hyphen as a word, while WORDS 2.0 and Word 6.0 do not. Since most of our calculations use the total number of words as a fixed denominator and this denominator tends to be quite large, slight increases or decreases have little effect on the overall analysis. Be aware, however, that these differences do exist—and are rarely documented.

6. In WordPerfect 6.0, this is found under Format/Line. In Word 6.0, it's located under File/Page Setup/Layout.