

# Language as research data: Application of computer content analysis in nursing research

Statements from clients are valuable research data in nursing. Traditional analysis of such data can be labor intensive and difficult to code reliably or to adapt to group comparisons. The Minnesota Contextual Content Analysis, a computer-assisted analysis program, offers a systematic approach to categorizing and reducing data and to interpreting manifest and latent meaning in linguistic communications. The text is used in entirety, and reliability of coding is ensured, even with great volumes of data and multiple variables. Social context of the communication and emphasized ideas are scored variables, allowing for statistical procedures to further refine the interpretation of meaning. This rigorous method can aid nurse researchers in using language as research data.

**Ann Wilde Kelly, PhD, RN**  
*Postdoctoral Associate*  
*Mayo Center for Nursing*  
*Mayo Clinic*  
*Rochester, Minnesota*

**A. Marilyn Sime, PhD, RN**  
*Professor*  
*School of Nursing*  
*University of Minnesota*  
*Minneapolis, Minnesota*

**L**ANGUAGE provides access to the meaning of human experience. The words people use reveal their construction of reality.<sup>1</sup> People express meaning in language that is most familiar to them; thus language use varies considerably. To serve as research data, such variety of use requires a systematic way to organize the analysis of language.

Research in nursing is oriented toward the understanding of human health experience. A growing body of nursing research explores clients' quality of life,<sup>2</sup> interpersonal relationships,<sup>3</sup> the meaning of health,<sup>4</sup> and other research issues about human, lived experience.<sup>5-8</sup> The data that reveal information about the meaning of experience are largely verbal, resulting from open-ended interviews.<sup>9,10</sup> The traditional methodologies associated with analyzing verbal data, however, lack quantification and are time consuming, labor intensive, often unreliable,

---

*The authors thank Donald G. McTavish, PhD, Department of Sociology, University of Minnesota, Minneapolis, for review of and comments on this article in its formative stages.*

*Adv Nurs Sci* 1990;12(3):32-40  
©1990 Aspen Publishers, Inc.

and difficult to adapt to group comparison. A method that retains the richness and sensitivity of interview data, while coding reliably and simplifying the analysis process, would make a significant contribution to nursing research.

A computer-assisted content analysis program, the Minnesota Contextual Content Analysis (MCCA), has been designed with the aim of eliminating some of the above problems with traditional analyses.<sup>11</sup> The method was used successfully in the nursing research project described in this article and holds promise for other research relying on interview data.

## CONTENT ANALYSIS OF TEXT DATA

Content analysis is the use of repeatable and valid methods for making specific inferences from text data to their content.<sup>12</sup> It is a technique that combines quantitative and qualitative analyses within a theoretic framework. With interview data, both quantitative and qualitative methods facilitate the understanding of the manifest and latent meaning inherent in the data.

Content analysis has an empirical orientation. It is exploratory, concerned with real phenomena, and predictive in intent.<sup>12</sup> Historically the roots of content analysis began with humans' conscious use of symbols and language. Content analysis has evolved as a process for analyzing data as symbolic entities in a contextual setting for a particular purpose.

Content analysis differs from other quantitative methods in its use of unstructured verbal material as data in its original form and in its entirety. In addition, it is distinct in being context sensitive, so that the data are not separated from the symbolic meaning they have for the subject.<sup>12</sup>

The central idea in content analysis is to classify the numerous words of the text into a few content categories. The key is choosing a strategy for reducing the information by extracting interesting and theoretically useful generalizations.<sup>12-15</sup> By supplying theory as the basis for the research, the researcher imposes his or her purpose on how the message source is delineated and on how messages are analyzed. It is the structure of the text, however, not the researcher, that should provide the objective references for content analytic methods.<sup>12</sup>

## Data reduction

Data reduction is achieved by dividing the text into units of meaning (idea categories) and quantifying these units according to rules.<sup>13</sup> A crucial aspect of content analysis is the construction of idea categories and the selection of words that are considered to be subsets of those categories. With traditional content analysis, the researcher designates idea categories from the underlying theory or ideas that are identified in the text. A reliability test must be undertaken if categories are formed in this way to show that another researcher has obtained the same result from the same rules and data.

One approach to categorization used in computer-assisted content analysis is the use of a general dictionary. Using a general dictionary has the advantage of standardization, which encourages comparable results across studies.

General dictionaries used in content analysis consist of numerous categories into which most words in most texts can be categorized. The categories of meaning are generally based on common sense and reflect the broad range of human experience and understanding that is encoded in language. If ambiguously defined, the categories can

prove a source of error. To avoid ambiguity, the categories are orthogonal. There are additional rules that distinguish among the various meanings of homographs according to the context of usage, a feature that has further helped to eliminate error.<sup>13,16</sup>

The computer, by virtue of handling highly complex coding, can analyze a text using more than one categorizing scheme. As long as the categories are mutually exclusive categories, validity is maximized.<sup>13</sup>

### **Interpretation**

The search for the latent meaning of the text is the interpretation stage in the analysis. This is considered a separate task from data reduction; it might be thought of as a form of model building. The text is a reality for which the textual analysis serves as a model.<sup>15</sup> The text supplies direct evidence of meaning, but it cannot be analyzed in the same way as something that can be directly observed. The meaning may be inferred by more or less rigorous procedures.<sup>12</sup> The relation among ideas, reality, and language is interactive.<sup>17</sup>

Interpreting the latent meaning of the data may be approached in various ways. A hermeneutic strategy is well suited to such interpretation. The computer facilitates the interpretation by displaying the data in visual formats, such as the distance between ideas or the clustering of ideas. In addition, the computer printout allows an emphasized idea to be traced to every instance of use without reviewing the entire transcript.

### **THE MCCA PROGRAM**

The MCCA describes and displays the ideas emphasized in a text, the relationships among the emphasized ideas, and the contextual orientation (social perspective) of a

text. In addition, the program compares and contrasts those three elements of the text among groups or individuals.<sup>11</sup>

The MCCA program was developed at the University of Minnesota and operates on a Control Data Cyber 174 computer. The program is currently being adapted for use with a personal computer.<sup>11</sup>

### **Idea emphasis score**

The dictionary used in this program is a general one oriented toward more frequently used words. The words are categorized by meaning into 116 idea categories that are mutually exclusive and that include about 90% of general word usage.<sup>11</sup> Those words that have several meanings are disambiguated by a scheme that takes into account the contextual perspective of the communicator.

In unstructured communication, there is a relation between the frequency of the ideas expressed and their saliency to the person speaking. The scoring of the emphasized ideas reflects this relationship. Scoring is achieved by a computer match of each word in the text against the words (indicator words) within the idea categories in the program dictionary. A tally of word usage is kept for each category. The percentage of total number of indicator words for each idea category in the text is derived from the total number of words in the entire text. The idea emphasis score is obtained by subtracting the percentage of overall usage for each idea category from an expected score obtained from a normed criterion. The resulting figure is divided by the variability in the use of idea categories across social contexts. A score of zero for an idea emphasis score (E score) on an idea category signifies that the idea was emphasized the same amount as one would find in a normed sampling of general written text. A plus or minus score indicates that the

idea was emphasized more or less, respectively, than the norm.<sup>11</sup>

The norms for use of written words, which form the standard for scoring in the MCCA program, are the expected probability of use of words in an idea category based on the probability distributions over a large sampling of public and written materials produced in the United States. The norms have been confirmed by previous work.<sup>11</sup>

Words that are not classified and scored by the program are identified separately on a printout. In most instances these are technical or misspelled words. Technical words may be assigned to an idea category, and misspellings are corrected before the computer run is repeated.

#### **Social context score**

The MCCA program also identifies and scores the social perspective or context from which the participant is communicating. The context is a socially defined structure or situation such as a social group, institution, or organizational culture. Membership in the group lends a particular cognitive perspective, and there is shared understanding of the boundaries of the communication among group members. The context is significant because the meaning of the communication is inseparable from it.

Four social contexts have been empirically identified in MCCA: traditional, practical, emotional, and analytic. These contexts are general in nature and orthogonal.<sup>11</sup>

#### **Traditional context**

The traditional context is a perspective emphasizing standards, codes, and rules that guide human behavior. Groups that take this perspective include religious, governmental, judicial, military, and familial. These groups focus on the establishment of beliefs,

norms, rules for behavior, and sanctions for role performance. For individuals, this perspective includes ideas of ought, should, and being "good" in a normative sense. The actors might be described as members, citizens, supporters, or followers. The words that are used suggest a concern for belonging to a system and conforming to norms within that system.<sup>11</sup>

#### **Practical context**

The practical context is one that focuses on the pragmatic aspect of a situation and goal achievement. Individuals displaying this orientation tend to be motivated by succeeding, achieving, getting ahead, controlling, and avoiding failure. A member of this group is often referred to by an occupational title such as manager, worker, or employee.<sup>11</sup>

#### **Emotional context**

The emotional context is one in which the individual thinks in terms of involvement, personal concern, comfort, and emotional expression. It is a reflection of ideas close to the self and motivations from personal experience. Illustrations of this perspective include the arts, music, poetry, organized recreation, and situations involving personal relationships. Sanctions often include comfort or discomfort, internal agony and strain, and satisfaction or release.<sup>11</sup>

#### **Analytic context**

The analytic context is expressed in intellectual and objective terms. Individuals express motivations of awareness, interest, understanding, knowledge, and curiosity. They achieve those through inquiry, evaluation, reflection, and analysis. This viewpoint reflects that of scientists, teachers, analysts, consultants, and researchers.<sup>11</sup>

The MCCA systematically codes contextual information directly from the communication rather than determining the context intuitively or using information outside the communication itself. The context is identified by the relative emphasis and vocabulary of the language. As each word is classified into an idea category, the relative use of each idea category is used as a factor to create a weighted score in each of the four general social contexts. The accumulated contextual (C) scores over a text are normed and are expressed as a positive or negative number. The scores on the four perspectives for any group (or individual) indicate the relative strength of each social perspective for the group (or individual).

### Interpreting the meaning of the language

A number of different approaches are used in the MCCA analysis to interpret the meaning of the language used in the text. Some of the most useful include identifying, in context, the specific words used in the idea categories; illustrating the differences among the ideas emphasized by groups; and displaying the ideas that cluster together.

### A STUDY OF THE MEANING OF HEALTH

Representation of health status from the perspective of persons in the health care system was studied. The purpose of the study was to explore with three groups of participants the kinds of representations they have of their health status and the salient features of the representations.<sup>18</sup> The study will not be reported in full. A few selected findings will be presented to illustrate use of the MCCA program.

The theoretic basis for the study came from a theory of self-regulation.<sup>19</sup> The theory

proposes that persons' representations of the threats to their health control their health-regulating behavior. Previous research exploring representations of threats to health was based on structured interview questions formulated by that researcher.<sup>20-22</sup> In the study reported here, participants' representations of their health status were elicited by a focused but unstructured interview approach.

Subjects for the study comprised three groups. One group consisted of 42 persons with hypertension (hypertensive group); another group of 8 persons with serious, chronic health problems without available treatment (untreatable group); and a third group of 16 persons with no acute or serious, chronic illness (normal group). The interviews resulted in 75,000 words of data. A content analysis was done using the MCCA.

### C scores, E scores, and interpretations

The context scores (Table 1) showed that all the groups had similar contextual perspectives. The positive emotional context scores indicate that all were communicating their personal experience.

The idea emphasis scores revealed those idea categories that were stressed by each group. Table 2 illustrates the scoring by group for a few emphasized ideas. Because the scoring indicates negative as well as

Table 1. Distribution of contextual scores (C scores) by group

Group	Contextual scores			
	Emotional	Practical	Traditional	Analytic
Normal	25	-12	-1	-12
Untreatable	25	-12	-3	-13
Hypertensive	25	-12	-1	-12

**Table 2.** Selected emphasized idea categories and idea emphasis scores (E scores) by group

Group	E scores			
	Depressed	Good	Object	Prohibit
Normal	1	16	9	-2
Untreatable	3	5	13	0
Hypertensive	0	15	12	-4

positive E scores, it was possible to identify ideas that the subjects were avoiding as well as those that they were emphasizing. The untreatable and hypertensive groups emphasized the "object" category, while the normal and hypertensive groups emphasized the "good" category.

To understand the meaning communicated by these emphasized ideas, the particular words used in each idea category and the relationships among the emphasized ideas were examined. This step constitutes the interpretation phase of analysis.

For example, interpreting one theme emphasized by the untreatable group proceeded as follows. The pattern of use for the group showed a high percentage of co-occurrence of four idea categories, that is, when one of the idea categories was emphasized, it was followed by one or another of three idea categories much of the time. The ideas were, "want responding," "begin action," "agony," and "depressed." The meaning of each idea was explored individually.

For example, the idea category "want responding" is defined by the program dictionary by a number of indicator words. The untreatable participants used the words *hope*, *need*, *needed*, *want*, *wanted*, *wants*, *will*, *wish*, *worked*, and *working*. The program provided a listing of all instances of the

use of each indicator word embedded in the text in which each was used. Examples of the indicators, in their context for the category entitled "want responding" follow.

They don't give out *hope* anymore.  
You *need* to be constantly exposed to positive stimulation.  
What people *want* is to be average.  
How disabled *will* I become before I die?  
It is hard; I *wish* I could cut down.  
Infection in my system, and they *worked* on it for about a year.

The second related idea, "begin action," was indicated by the words *begin*, *beginning*, *encouraging*, *risk*, *seek*, *start*, and *started*. Examples of each in their context follow.

In the *beginning* you go through several responses.  
They are always *encouraging* positive emotions.  
Increased *risk* of infection perhaps.  
The ones who don't *seek* support are afraid.  
I'll never run again so I have to *start* there.

The third co-occurring idea category, "agony," was indicated by the words *pressure*, *trouble*, and *illness*. Examples of their use follow.

Take off the *pressure*, the pressure of whatever you're doing.  
A great deal of the bronchial *trouble*. . .  
You go through several stages of response to the *illness*.

The fourth idea, "depressed," was indicated by the words *sad*, *tired*, *worried*, and *depressed*. The following are examples of how the indicator words were used in the text.

You realize you may die from this and you get *depressed*.  
I get *tired* and awfully achy.  
I did think about the risks. One thing I was *worried* about was head involvement.

The act of interpreting the themes required a reflective procedure beyond inductive or deductive reasoning.<sup>23-27</sup> A hermeneutic approach was used that involved a dialectic movement, a scanning of the data from part to whole and back to part again. A flow of meaning expressed as themes was then abstracted from the text. The limited number of examples from the text provided here does

not reflect the meaning of the ideas as fully as is possible when all instances of use in the complete text are reviewed. The latter case results in a full and rich sense of meaning. The interpreted representations were derivations of the experienced meaning. In addition, a peer review supported the conclusions drawn by the researchers.

The interpretation of the above relationship of ideas was that the untreatable group wanted a certain health state but accepted that the outcome was uncertain. They accepted that a health problem existed and recognized that change in behavior was necessary to respond to the health problem and to their own emotional response. Sadness and worry persisted for this group.

### Comparing groups

Because this study compared the ideas expressed by the three groups, cluster analysis, another element of the program, was useful in determining whether each group was distinct with respect to its emphasis of ideas. When other clusters of idea categories were analyzed, a clear separation of groups was shown.

Scoring allowed statistical procedures to be applied to the data. In this project a factor analysis assisted in further defining the themes of the representations of health—both those themes held in common by all participants and those that were unique by group. In addition, a discriminant analysis further identified the differences among the groups.

It was found that 11 themes about health were emphasized by all participants. The salient themes for the subjects included the experience of their health or illness in the context of their own life, their distress resulting from their health status, and overall

changes in their life and personal roles because of their health status.

Each group of subjects also emphasized unique themes. The untreatable group emphasized controlling the distress about their condition, limiting their focus on the disease and its outcome, and strengthening their inner spirits. The hypertensive group stressed the concrete actions they were taking to control the high blood pressure, their support systems, and their reliance on others for help in management of their health. They seemed generally optimistic about the outcome of their condition. The normal group stressed uncertainty about their future health and talked about health-promoting actions that they were doing or should be doing. They expressed difficulty in carrying out health-promoting activities. A detailed discussion of the findings may be found elsewhere.<sup>18</sup>

### USING THE MCCA

Operating the MCCA programs was relatively simple. The voice-recorded interviews were transcribed and uploaded to a mainframe Cyber for running the program. A normal format was used with simple codes inserted at the beginning and end of each segment of the text to be contrasted. In the case of this study, one run contrasted individual interviews, and the next run contrasted the three groups of interviews. The coding had to be changed between runs.

The computer-assisted content analysis allowed categorization of the entire text so that no contributions to meaning were omitted. The preconstructed categories ensured their independence and objectivity, scoring permitted the statistical analyses, and visual displays provided a fresh vision of the relationship of ideas to the text.

• • •

The MCCA is an innovative methodology that offers a rigorous analysis of language. It is a method useful for analyzing contextual orientation as well as the manifest and latent meanings inherent in linguistic expressions. The program preserves the integrity of the data, ensures the reliability of coding, and maintains sensitivity to the meaning of the language. MCCA has the capacity to process great quantities of data and to deal with multiple variables beyond the scope possible with human coders. The perfect reliability of coding allows researchers to replicate studies by using available data banks. By displaying the data

in a number of ways, the researcher has access to different levels of meaning. Systematic differences among individuals or groups can be identified by repeating the computer analysis using different comparisons. By integrating the quantitative and qualitative analyses, MCCA achieves a comprehensive approach to analyzing linguistic meaning and offers a contribution to the repertoire of methods available for analyzing verbal data.

Other computer content analysis programs undoubtedly will be developed. Such powerful and useful programs will surely aid nurse researchers in the use of language as research data.

## REFERENCES

1. Leonard VW. A Heideggerian phenomenologic perspective on the concept of the person. *ANS*. 1989;11(4):40-55.
2. Benner P. Quality of life: a phenomenological perspective on explanation, prediction and understanding in nursing science. *ANS*. 1985;8(1):1-14.
3. Powers BA. Social networks, social support, and elderly institutionalized people. *ANS*. 1988;10(2):40-58.
4. Moch SD. Health within illness: conceptual evolution and practice possibilities. *ANS*. 1989;11(4):23-31.
5. Bowers BJ. Intergenerational caregiving: adult caregivers and their aging parents. *ANS*. 1987;9(2):20-31.
6. Olshansky EF. Identity of self as infertile: an example of theory-generating research. *ANS*. 1987;9(2):54-63.
7. Rigdon IS, Clayton BC, Dimond M. Toward a theory of helpfulness for the elderly bereaved: an invitation to a new life. *ANS*. 1987;9(2):32-43.
8. Haase JE. Components of courage in chronically ill adolescents: a phenomenological study. *ANS*. 1987;9(2):64-80.
9. Sandelowski M. The problem of rigor in qualitative research. *ANS*. 1986;8(3):27-37.
10. Giorgi A. Theoretical justification for the use of descriptions in psychological research. In: Ashworth PD, Giorgi AJ, de Koning AJJ, eds. *Qualitative Research in Psychology*. Pittsburgh, Pa: Duquesne University Press; 1986:3-22.
11. McTavish DG, Pirro EB. Contextual content analysis. *Quality Quantity*. In press.
12. Krippendorff K. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, Calif: Sage; 1980.
13. Weber RP. *Basic Content Analysis*. Beverly Hills, Calif: Sage; 1985.
14. Woodrum E. Mainstreaming content analysis in social science: methodological advantages, obstacles, and solutions. *Soc Sci Res*. 1984;13:1-9.
15. Lundkvist K. Approaches to textual analysis. In: Rosengren KE, ed. *Advances in Content Analysis*. Beverly Hills, Calif: Sage; 1981.
16. Markoff J, Shapiro G, Weitman SR. Toward the integration of content analysis and general methodology. In: Heise DR, ed. *Sociological Methodology: 1975*. San Francisco, Calif: Jossey-Bass; 1975:1-58.
17. Deese J. Conceptual categories in the study of content. In: Gerbner G, Holsti OR, Krippendorff K, Paisley WJ, Stone PJ, eds. *The Analysis of Communication Contents: Developments in Scientific Theories and Computer Techniques*. New York, NY: Wiley; 1969.
18. Kelly AW. *Representations of Health in Persons with Chronic Health Problems*. Minneapolis, Minn: University of Minnesota; 1989. Dissertation.
19. Leventhal H, Johnson JE. Laboratory and field experimentation: development of a theory of self-regulation. In: Woolridge P, Schmitt MH, Skipper JK Jr, Leonard RC, eds. *Behavior Science and Nursing Theory*. St Louis, Mo: Mosby; 1983:189-262.
20. Meyer D, Leventhal H, Gutmann M. Common-sense



- models of illness: the example of hypertension. *Health Psychol.* 1985;4(2):115-135.
21. Ringler KE. *Process of Coping with Cancer Chemotherapy*. Madison, Wis: University of Wisconsin; 1981. Dissertation.
22. Turk DC, Rudy TD, Salovey P. Implicit models of illness. *J Behav Med.* 1986;9(5):453-474.
23. Polkinghorne D. *Methodology for the Human Sciences: Systems of Inquiry*. Albany, NY: State University of New York Press; 1983.
24. Geertz C. *Local Knowledge*. New York, NY: Basic Books; 1983.
25. Wilde D. *Hermeneutic Phenomenology: The Philosophy of Paul Ricoeur*. Evanston, Ill: Northwestern University Press; 1971.
26. Klemm DE. *The Hermeneutical Theory of Paul Ricoeur: A Constructive Analysis*. Toronto, Canada: Associated University Press; 1983.
27. Spiegelberg H. *Phenomenology in Psychology and Psychiatry: A Historical Introduction*. Evanston, Ill: Northwestern University Press; 1960.