

MAPPING STRATEGIC THOUGHT

Edited by

Anne Sigismund Huff

University of Illinois at Urbana-Champaign

JOHN WILEY AND SONS

Chichester · New York · Brisbane · Toronto · Singapore

11

Content Analysis

Carolyn B. Erdener and Craig P. Dunn

Baylor University and Indiana University

Overview

Content analysis as a research method for analyzing written communication has been employed extensively in the social sciences, foremost perhaps in political science. In recent years it has seen increasing application in the fields of marketing and to a lesser extent management. Although content analysis has not been the method of choice for the majority of management researchers, a small number of research projects based on this method has been reported in scholarly management journals (Bettman and Weitz, 1983; Dirsmith and Covaleski, 1983; Pearce and Davis, 1987; Salancik and Meindl, 1984; Ricchio and Belohlav, 1983). The primary focus in all these articles is not on the method *per se*, but rather on an interesting and significant research question that could be effectively addressed by means of its application. When applied to the study of important issues and theoretical questions, content analysis can yield insights that would be difficult to derive from other methods.

Although content analysis has become a familiar and generally well-accepted tool for research in the social sciences, its use in the field of management appears to have been somewhat restricted by problems of execution. These include the tedious, inordinately time-consuming process of carrying out content analysis manually, and

the inherent problems of human error and subjectivity of interpretation which must be addressed. These problems are all the more critical where large quantities of text are to be analyzed. A substantial portion of the current chapter is devoted to explication of possible technical mechanisms for resolution of such problems.

It is our position that recent technological developments have considerably reduced the problems associated with performing content analysis. This reduction in turn increases the potential usefulness of this method for organizational researchers. Important developments include the variety and sophistication of software packages currently available for computerized content analysis, and in particular the recent adaptation of some of these for use on personal computers. Specifically, two extant software programs representing very different approaches to the analysis of digitized textual material will be examined—Textpack V and Logic-Line Series 2. Equally important are significant improvements in the range (in terms of both cost and capabilities) of optical character recognition scanning equipment now on the market.

Usefulness

Content analysis has a number of strengths and limitations which should temper the researcher's decision as to when and how to use this method. Among the advantages of content analysis in general are:

1. it permits systematic interpretation of textual material based on objective criteria;
2. it can facilitate the distillation of large quantities of information down to a manageable size;
3. it can convert qualitative material to quantitative data, which then become the basis for further analysis using standard statistical procedures;
4. it does not necessarily require large amounts of data, but can also be used for small-scale studies that do not impose heavy resource commitments on the researcher;
5. it is the ultimate unobtrusive measure; and
6. it can be used in combination with other research methods—such as the quantitative analysis of financial statements, or of survey data collected in the field—to combine fine-grained with coarse-grained research methods in triangulating on complex issues or organization and strategy (Harrigan, 1983).

The primary limitation in the use of this research method revolves around an inherent trade-off between validity and reliability. As Anne Huff has so aptly pointed out in the introductory chapter, the central benefit of the cognitive map is that it encourages holistic synthesis rather than reductive analysis. This principal advantage is, however, largely eroded if one employs content analysis in one of its more traditional, deductive modes. Conversely, if content analysis is used to argue from the

parts to the whole—the inductive approach—the 'real' meaning behind the written word(s) emerges, thereby engendering holistic synthesis. This distinction has a clear correlation in the fact that content analysis can be conducted on one of two levels: (1) *manifest content*, which captures various surface characteristics of the words used (Berelson, 1952); and (2) *latent content*, which captures the underlying meaning embodied in the text.

The analysis of *manifest content* focusses on such features as word frequency counts (the frequency of occurrence of predetermined key words) and key words in context (the relation of the key word to other words in the sentence, in terms of syntax and semantics). This can be measured in terms of: a raw score; a percent of total words; or a ratio in comparison with the occurrence of other sets of words in the same text (or of the same set of words, in other texts). The critical assumption is that these key words relate to underlying concepts or constructs that are germane to the research question at hand; however, there is rarely if ever a perfect match, and so the validity of this measure is always subject to challenge. For example, content analysis of annual reports based on word frequency counts or the analysis of key words in context using words related to new product development will not capture textual material which identifies and discusses new products by brand name without mentioning the key words (e.g. 'new product,' 'product development').

The analysis of *latent content*—the underlying, deep meaning embodied in a text—better addresses the researcher's concern for validity. However, since it depends to a greater extent on subjective interpretation and judgement, latent content analysis introduces a tradeoff in terms of reliability. This problem is usually addressed by means of additional precautionary measures to cross-check the coder's potentially subjective interpretation.

Procedure

Regarding comments in the introductory chapter relating to the 'territory' to be mapped, it will be recalled that the most basic decision involves consideration of the choice between a map that reflects individual cognition and one that reflects the shared perceptions of a group. It cannot be stressed strongly enough that the research question under analysis should inform this decision. For example, if the researcher's interest is in ascertaining the CEO's perspective of enterprise strategy, then content analysis might appropriately focus on the CEO's own business correspondence, public speeches, interviews, and perhaps annual letter to the shareholders. However, if the researcher's concern is with determining how financial institutions perceive the firm's strategy, the textual sample would necessarily include documents generated by their own analysts.

The major procedural decisions in performing content analysis revolve around the issues of coding and sampling text. General considerations are summarized below (Babbie, 1986; Holsti, 1969; Krippendorff, 1980; Webber, 1985). Coding consists of

first establishing a list of interpretive categories, and then ranking the text at hand in terms of its relevance to each category. It requires the rater to interpret the material according to these predefined categories. To guard against overly subjective interpretation and to enhance reliability, each text can be interpreted and coded by multiple raters (Bettman and Weitz, 1983; Pearce and Davis, 1987). It can also be coded and then recorded by the same rater after a significant period of time has passed (Enz, 1986). The objectivity of the coding can then be expressed quantitatively, in terms of inter-coder reliability coefficients.

Coding is particularly important for latent content analysis; however, it also has application in the process of initial identification of key words used in manifest content analysis. Ideally, the key words list should be taken from a standard content analysis 'dictionary' that has been widely tested and is generally available to other researchers, cf. in the social sciences the *Harvard IV Psychosocial Dictionary* and the *Lasswell Value Dictionary* (Webber, 1984, pp. 24–34). If no appropriate dictionary has been compiled, the researcher will have to create one, by identifying sets of words that relate to those concepts of interest to his/her particular research. This process could be carried out by manually coding a subset of the text to be content analyzed, as described above; by means of a general survey; or through use of an interactive computer program for text analysis, such as Logic-Line Series 2, described below.

The selection of texts to be analyzed involves sampling decisions. Once a general body of material relevant to the research question has been identified—for example annual reports of American, Japanese, and West German firms—some decision rule must be used to select an appropriate textual sample. This decision rule should be formulated with the unit of observation and the unit of analysis clearly in mind (e.g. annual reports as the units of observation [or, more specifically, letter to shareholders, table of contents, subtitles], and individual companies as the unit of analysis). Conventional sampling procedures—whether random, systematic, or stratified—can then be applied. In the scenario under review, random sampling would entail selecting a variety of annual reports from each country in question, irrespective of other considerations such as industry or year. Alternatively, systematic sampling would require selecting annual reports according to some pattern, such as every tenth item in the list of possibilities, or the first company under each letter of the alphabet in the card catalogue of the library in which the reports are stored. If the scope of the analysis were to include the body of the text of each annual report selected, random or systematic sampling could be used to narrow the scope of the analysis to, for example, the first complete sentence on every page, or every fifth paragraph. This is of particular interest when content analysis is performed manually; it is less an issue for computerized content analysis, which has the capability to easily process large quantities of text. Finally, stratified sampling would entail classification of annual reports in terms of such theoretically relevant variables as revenues, profits, or number of employees. Cluster sampling can be used with any of the above techniques to maximize the advantages of availability and accessibility.

Textpack V¹

The introductory chapter to this volume included a discussion of the five cartographic choices for mapping strategic thought. This continuum provides a useful vehicle for categorizing both Textpack V and Logic-Line Series 2 (the latter to be discussed in the following section). At one end of the continuum are those techniques which assume the underlying model of cognitive activity to be relatively simple—that is, cognitive importance is deduced from word usage. Textpack V clearly falls within this category.

Textpack V is designed for quantitative content analysis as well as data management in qualitative text analyses. This program package includes subprograms for word frequency counts and concordance suitable both for analysis with a strict dictionary approach (using a predetermined content analysis dictionary) and for empirical approaches involving word clustering. Texts are normally transferred into Textpack V through the media of punch cards, optical character reader (OCR) or direct data entry. If a Kurzweil data-entry machine or other OCR is used for data entry, the next step is the correction of the text file. It is recommended that such data 'clean-up' be completed in one of the more popular text-editing programs (e.g. WORDSTAR, for personal computers, or XEDIT, for IBM mainframe computers), as Textpack V contains only a rudimentary editor program. A resident spell-check program proves to be of immense help during this process.

In general, Textpack V defines a word as a sequence of characters set apart by one or more blanks. For the purpose of discrete word identification, therefore, the Textpack V user is encouraged to purge text of punctuation through use of the STOP LIST subprogram. Character strings are limited to thirty-nine consecutive characters. A word root is defined as 'the left part of the word in the sense of the general word definition.' Therefore, utilization of a root word search results in selection of all words which begin with the root word specified, whether or not such root word is followed by a recognizable suffix (e.g. if a word search concerning the issue of *equitable distribution of organizational resources across stakeholder groups* was initiated predicated on the root word 'just,' a reference to 'Just-in-Time' might be as likely to turn up as one to 'justice'). Multiple word combinations may be specified, though the thirty-nine consecutive character limit still applies. In order to differentiate among homonyms, disambiguation (through 'coding' homonyms within the text as different words—'spring-noun,' 'spring-verb,' etc.) is required if the problem of word similarity is to be overcome.

All of the above occurs, of course, before the real analytic work of Textpack V begins. Textpack V provides for six basic manipulations:

¹This section is reprinted with permission of ZUMA (Zentrum Für Umfragen, Methoden und Analysen e.V.).

1. *Vocabulary with word frequencies*—a vocabulary is a simple means of describing a text . . . here it includes the frequencies for different words which are counted for one or more texts.
2. *Index*—an index is a list of words occurring in a text.
3. *Key words in context (concordance)*—the meanings of single words are frequently ambiguous, and thus it becomes necessary for semantic analyses to locate words in their contexts.
4. *Comparison of vocabularies*—designed to compare two vocabularies for common or differing terminology.
5. *Text selection using external information*—if the user has additional information pertaining to a text in the form of a numeric file, then he/she can . . . select sections from a text which correspond to a specific combination of values in the numeric file.
6. *Coding*—the user can record which words occur in specific text sections and store this information in a numeric file . . . if desired, he/she can then perform statistical manipulations with SPSS or OSIRIS on the basis of this information.

(Excerpted from instructional materials prepared by Peter Ph. Mohler and Cornalia Zuell of ZUMA* to accompany the Textpack V personal computer program for content analysis—included here with the authors' permission.)

Textpack V allows for specification of up to three levels of data analysis, which may then be run concurrently. In addition to their use as an aid in focussing any given search (through the Textpack V Filter IDn=m subprogram), identification of data levels (e.g. 'annual report,' 'shareholder letter,' and 'paragraph') allows for data transformation routines analogous to the 'SELECT IF' subroutines within SPSS.

By virtue of its reductionist orientation, Textpack V stands in contrast to the 'holistic synthesis' perspective illuminating the vast majority of this volume. In support of this observation, consider the closing words of the introduction to the Textpack V manual:

The TEXTPACK programs are actually little more than a kind of meat grinder: a piece goes in whole at the top and comes out below chopped up. And as with the meat grinder something is put in and something comes out: input and output. Technically speaking, both are files—a SENTENCE file, for example, which is put into the program LISTSPPLIT and which comes out as a SPLIT file. The user tells the program with parameters what the input and output is.

Logic-Line Series 2

At the opposite end of the cartographic spectrum is Logic-Line Series 2. The basic assumption underlying this approach is that language can be taken as a sign of underlying structure. While necessitating an admittedly interpretative posture, this perspective has the advantage of bringing the notion that cognition is highly

conditioned by previous experience to bear upon the research question. Logic-Line Series 2 is specifically designed to use the logic of 'fuzzy sets' to illuminate the holistic structure underlying textual evidence.

Logic-Line Series 2 has the ability to retrieve straight symbolic references, numeric patterns, and combinations of text and numerics from a larger body of textual material. Additionally, the program is capable of drawing cross-correlate inferences from both textual and numeric references. While Logic-Line Series 2 has the means to search, study, and retrieve information, what distinguishes this package from other content analysis programs is its ability to engage in intelligent cross-association algorithms. In this mode the program actually evidences a form of artificial intelligence or, more specifically, an expert system shell.

In principle, this program package is deceptively simple. Logic-Line Series 2 operates by assigning a numeric code to each key word in both the search phrase and the text file, and then searching for matching patterns—which are then automatically retrieved as a point of inference. The logic of fuzzy set theory allows for consideration of 'close' matches. At the point of retrieval, the researcher has the option of verifying such a 'match.' This information is then used by the program as input information for the purpose of subsequent trials. It is this feature which allows for 'learning' (or, in a stricter sense, an increase in the program's familiarity with text files) within the machine environment.

It should be apparent from the above discussion that Logic-Line Series 2 is extremely dependent upon the researcher. Such dependence was a matter of rational choice on the part of the software's developer. The program is not designed to be deterministic, or even probabilistic, in nature; rather, possibilistic reasoning is utilized as a means of mirroring subcognitive thought processes. Logic-Line Series 2 is therefore not appropriately thought of as a 'black box' through which inputs are mysteriously transformed into outputs, but as an assistive device for the researcher engaged in, for example, holistic cognitive mapping. The product of the program might therefore be construed as accelerated mental functioning rather than actual data.

The possibilities for employing Logic-Line Series 2 within organizational research are limited only by the researchers' imagination. While the program can replicate conventional quantitative content analysis with greater efficiency than programs such as Textpack V, it has the additional capability of correlating textual and financial analysis within, for example, examination of corporate annual reports. In a similar manner, the program might be used to compare the results of programmed survey results with those of open-ended questionnaire responses within the same research program.

Additionally, Logic-Line Series 2 may be used to develop a dictionary of terms for utilization within the more traditional quantitative textual analysis programs. Consider for the moment a research project concerned with references to 'research and development' within annual reports. Standard quantitative content analysis programs are likely to uncover only a limited number of such references, while in the search process overlooking a host of other references which refer to research and development

without using those exact terms. A relatively small sample of annual reports might be analyzed using Logic-Line Series 2, resulting in a valid dictionary of terms relating to the construct of research and development. This dictionary file could then be directly employed as the search file within a straightforward quantitative content analysis program—thus saving time over using Logic-Line Series 2 for the entire search, while commensurately allowing for extraction of a greater number of relevant textual references.

Similar reasoning would allow the use of Logic-Line Series 2 for specification of operational measures for a given construct in cases of, for example, survey instrument development. It is at this juncture that the possibilities become truly manifest. Given a developmental survey instrument, the researcher might pre-test the instrument in the laboratory. Individuals representative of the respondent profile could be enlisted to perform content analysis as a means of insuring that the appropriate constructs are being tapped by the questionnaire. As has been noted previously, Logic-Line Series 2 is fundamentally an inductive approach; therefore, sample subjects ought to be able to discover the constructs underlying individual (or groups of individual) survey items—and thereby (dis)affirm the reliability of the survey instrument prior to final administration.

Subject interaction with the program itself is not limited, however, to survey development. The program package might be used as a means of assessing the consistency of organizational communications across various dimensions. It might be discovered, for example, that a corporation's code of ethics has vastly different meanings depending upon one's level within the organization. Similarly, whether one finds the same code of conduct to be wholly (in)adequate might be dependent upon which of the organization's stakeholder groups one considers to be one's primary organizational referent—or of which country one is a citizen, or which industry one represents, or any number of other independent considerations. It should be apparent that Logic-Line Series 2 has far-reaching implications for research concerned with mapping strategic thought.

Optical Character Recognition Scanners

No discussion of this methodology would be complete without at least a passing reference to optical character recognition (OCR) scanning equipment. It has been our experience that many researchers have an interest in utilizing content analysis as a means of mapping cognitive thought; however, the sheer tedium of data entry has in most cases made such efforts impracticable. Recent developments in both hardware and software have greatly increased the feasibility of doing such studies.

The newer programs are capable of recognizing a vast array of type styles in a wide variety of point sizes—and even hand-printed text! Additionally, such scanners are now able to handle multiple columns and integrated graphics with relative ease. This is not to suggest, however, that such hardware/software combinations operate with

complete accuracy. The typical micro-based system might have a 'hit' rate of 96 percent, which translates to about 120 missed characters per average page of 3000—though some vendors are now claiming accuracies as high as 98 to 99.5 percent. Although 120 missed characters per page might seem inordinately high, such textual information can be 'cleaned up' with relative ease in any of the more popular word-processing programs. Spell-checking programs are able to locate virtually all of the textual errors, and many have the additional capability of correcting consistently miscued words throughout entire documents with a few simple keystrokes.

Many academic institutions now operate dedicated optical scanning equipment. If the researcher finds such hardware/software to be unavailable, a host of companies offer microcomputer-based systems. These range in price from several hundred dollars to several thousand dollars depending upon features, not the least of which are speed and accuracy. These prices do not, of course, reflect the cost of the basic personal computer—which, incidentally, needs expansive random-access-memory (RAM) capability—used to run such applications. Most universities offer researchers the services of computer consultants who should be apprised of the latest in OCR technology. Additionally, many of the more popular computer journals offer review articles of such equipment. As OCR technology changes almost daily, the researcher interested in such a major purchase is well advised to seek the most current literature available.

Concluding Observations

As with all research tools, content analysis can result in interesting and significant findings when appropriately used and properly carried out. In the field of strategic management, it is perhaps most powerful in combination with other research methods. For example, it can be used to gain a different perspective on an issue that has been studied by other means, sometimes generating conflicting results. Interpretation and resolution of these discrepancies can trigger new insights, as in the case of the dual analysis of letters to shareholders and financial statements in corporate annual reports (Salancik and Meindl, 1984). Content analysis can also be employed as a middle step in generating data for statistical analysis.

Given the increasingly widespread availability of optical scanning equipment, data banks of relevant texts including annual reports, and personal computer programs for content analysis, we may expect to see more interest in the use of this research method.

References

- Babbie, Earl (1986) *The Practice of Social Research*, 4th edn. Belmont, CA: Wadsworth.
Berelson, B. (1952) *Content Analysis in Communications Research*. Glencoe, IL: Free Press.

- Bettman, J. R. and Weitz, B. A. (1983) Attributions in the boardroom: causal reasoning in corporate annual reports. *Administrative Science Quarterly*, **28**, 165–183.
- Dirsmith, M. W. and Covaleski, M. A. (1983) Strategy, external communication and environmental context. *Strategic Management Journal*, 137–151.
- Enz, Cathy A. (1986) *Power and Shared Values in the Corporate Culture*. Ann Arbor, Michigan: UMI Research Press.
- Fiol, C. M. (1985) Underlying the motivation to form a domestic joint venture. Presentation at Academy of Management National Meetings, San Diego, California.
- Harrigan, K. R. (1983) Research methodologies for contingency approaches to business strategy. *Academy of Management Review*, **8**, 398–405.
- Holsti, O. K. (1969) *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.
- Krippendorff, K. (1980) *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, CA: Sage.
- Pearce, John A. II and Davis, Fred (1987) Corporate mission statements: the bottom line. *Academy of Management Executive*, **1**, 109–116.
- Salancik, G. R. and Meindl, J. R. (1984) Corporate attributions as strategic illusions of management control. *Administrative Science Quarterly*, **29**, 238–254.
- Sussman, L., Ricchio, P. and Belohlav, J. (1983) Corporate speeches as a source of corporate values: an analysis across years, themes, and industries. *Strategic Management Journal*, **4**, 187–196.
- Webber, R. P. (1985) *Basic Content Analysis*, Beverly Hills, CA: Sage.

The Repertory Grid Technique for Eliciting the Content and Structure of Cognitive Constructive Systems

Rhonda K. Reger

Arizona State University

The repertory grid technique was first developed by George Kelly (1955) to operationalize his personal construct theory. The technique remains closely associated with Kelly's (1955) theory of cognition and appropriate use of the method requires familiarity with the theory and acceptance of its assumptions. Detailed discussion of the theory is beyond the scope of this chapter; interested readers should consult Kelly's original works.

In developing repertory grid technique, Kelly was interested in developing instruments in which the researcher's frame of reference and worldview would not be imposed on the respondent. At the same time, he needed a method that would reliably elicit the respondent's cognitive structure. Close-ended surveys were rejected because they impose the researcher's cognitive structure. Open-ended interviews were rejected since most people are not conscious of the ways they cognitively organize and could not give valid and reliable answers to direct, open-ended questions.

Repertory grid technique was developed to fill the void. It is a set of innovative interview techniques which elicit responses in a semistructured manner, but the content and specific structure of responses vary by respondent. Repertory grid data can be analyzed using a number of qualitative and quantitative analysis methods.