

A NOTE ON CONTENT ANALYSIS

ATIS K. DASGUPTA

Indian Statistical Institute, Calcutta

In India, application of content analysis as a research technique is a recent development in the field of social sciences. In 1967 a project was initiated in the Sociological Research Unit (SRU) of the Indian Statistical Institute to find out the possibility of introducing content analysis technique to appraise views on societal changes in the Indian context (Mukherjee *et al.*, 1971). In September 1974, the International Social Science Council of UNESCO invited this author from SRU to an international workshop in Pisa (Italy) where scientific workers doing research on content analysis participated from different countries of Europe and North America. Workshop of this kind has seldom taken place in the history of content analysis, and it has brought into focus all the major trends of this research technique currently in vogue. However, as I am not sure whether most of the social scientists in this country are already familiar with content analysis in general, a brief summary about the basic characteristics of this research technique and the history of its development may be found useful, before discussing the present position of content analysis and the problems connected with its recent applications.

BASIC FEATURES

Content analysis is generally defined as a 'research technique for making inferences by systematically and objectively identifying specified characteristics within text' (Stone *et al.*, 1966). Primary importance is given to the theory or hypothesis being investigated. The theory and hypothesis determine the texts to be compared, the characteristics to be specified for measurement from within the texts, and the kinds of inferences that might be drawn from the results of measurement. Upon selection of text, the content analysis process makes on it systematic and objective measurements by identifying occurrences of specified characteristics. These measurements then serve as a reference for drawing

inferences. While the measurement process is explicit and direct, the inferences drawn from the results of measurement may be implicit and indirect. Whether the measurements and inferences follow reasonably from the text may be ascertained by checking back to it.

Now, one can ask, 'where, then, is the difference between conventional research and content analysis technique?' One may point out that any research in social sciences involves some kind of measurement and inference. In conventional research inferences are primarily made from the careful inspection of the text itself rather than from the intermediary steps of explicit measurement. Each inference is made from the context of a particular event rather than from the measures abstracted by the content analysis process. In spite of this difference, however, the role of content analysis is not contradictory, but complementary, to conventional research approach. In both types of research, the investigator, to begin with, should acquaint himself with some knowledge about the nature of data relating to his theme of research. 'From this knowledge he should organise his hypothesis. At this point, in a conventional study, he would start writing. At this point, in content analysis, he is ready to set up his categories, to apply them to text, and then to start counting' (Lasswell *et al.*, 1952). This intermediary stage of explicit measurement is expected to make the basis of inference both systematic and objective.

EARLY HISTORY

The tradition of research labelled as content analysis had its beginning in the field of American journalism about the turn of this century: by the 1920's it was centred in the School of Journalism of Columbia and culminated in Malcolm Willey's study, *The Country Newspaper*, in 1926. Typically the early investigators used to measure, from newspapers and journals, attention devoted to straight topic categories, such as fashion, sports, and so on. The actual measures were usually taken in terms of the amount of space (such as column inches) devoted to the topics. A large portion of such studies had been mechanical and superficial tabulations of 'who says how much of what to whom', with no clear significance for any theory or hypothesis.

It was the work of Harold Lasswell and his associates in the 1940's that put content analysis technique in a much more refreshing and meaningful perspective. Based on an extensive study of political propaganda,

their work offered a major opportunity to make advances both in conceptualization and technique, summarised in their book *Language of Politics* (Lasswell *et al.*, 1949).

DEPARTURE FROM EARLY PRACTICE

From the middle of 1950's application of content analysis increased not only in journalism and political science but also in other fields, like—psychology, history, anthropology, education, sociology, linguistics, philology and literary analysis. However, by the beginning of 1960's it was realised that it had generally been difficult, expensive, and time consuming to employ human coders to follow complex content analysis directions dealing with large amounts of textual material. Individual differences in actual coding practices often occurred and the same coder showed considerable fluctuations over time. These considerations prompted the content analysis to turn to computer which opened up a new phase in content analytical research. It was envisaged that the computer could take over the tedious aspects of handling textual material of large volume with a speed unthinkable in the case of human coders. It was further suggested that the implementation of procedures on a computer would ensure that these would be both systematic and objective. The nature of a computer programme would require that both the categories relevant to content analysis and the rules for identifying them occurring in the text should be explicitly stated. The computer would then systematically apply the specified categories and the rules to the textual data in a completely objective manner. Whether the measurement procedure would be reasonable or best suited to the inferences being made might be debated, but the procedure itself would be explicit and unfluctuating. However, the presence of a computer serves to dramatize a choice. It can be a reliable aid to the social scientist. At the same time, the availability of a powerful computer makes it tempting to forget theory and hypothesis and simply let the method tell the story. Any way, before going to the controversies connected with the computer-aided content analysis, let us first consider how automated analysis technique actually works.

USE OF COMPUTER

For this purpose, we have to turn to Professor Philip Stone of Harvard University who started this kind of research (Stone *et al.*, 1966). In 1966

Stone and his associates built up a set of programmes called General Inquirer to introduce the operation of computer-aided content analysis. In order to explain what this set of programmes is about, we can start with recollecting once again the generally agreed definition of content analysis: A research technique for making inferences by systematically and objectively identifying specified characteristics within text. The 'specified characteristics' referred to may be represented as a system of categories and such systems may be of various sorts; as for instance, a simple list of specified words or phrases to be identified and counted or thematic categories derived from social science theories.

'Objective and systematic' identification of such specified characteristics or categories implies the need for a comprehensive and explicit source of reference against which assignments may be checked. This may take the form of a listing of all possible textual occurrences, against which a particular textual occurrence will be checked so that the appropriate assignment may be made. This procedure depends on the formulation of a special type of dictionary which will contain two basic types of information: (i) a list of all possible textual items on which category assignments may be made, and (ii) a specification of the categories to be assigned in each case.

In addition to the construction of specialised dictionary, the working of automated content analysis involves four major functions:

- (a) the preparation of data which implies the transference of text, following certain minimal rules of format, to punch cards;
- (b) the formulation of a programme which scans the transferred text sentence by sentence and looks up in the dictionary for each word it encounters and applies the appropriate categories (where a word has more than one possible sense the programme goes through a series of predetermined disambiguation rules);
- (c) the formulation of a separate programme which produces a print-out showing, among other things, (i) a word-by-word, sentence-by-sentence, list of all category assignments, and (ii) a list showing the number of occurrences of each category and the percentage of words and sentences in which each category appears; and,
- (d) the construction of a set of retrieval programmes to help the researcher for 'making inferences'. Once the measurement process is completed, the investigator is likely to search for 'themes' with-

in analysed text—for certain co-occurrence of categories of words, phrases, etc. The retrieval programmes are designed to assist the investigator in this type of inference process.

Before ending with this brief summary of automated content analysis, we should refer to one latest technological innovation which replaces the card punching operation by a specialised micro-film technique acceptable to large-memory computers. This is known as optical reader system which is capable of processing text printed in a variety of typeset fronts; these machines are very speedy and now have an error rate considerably below that of an average key puncher.

CRITICISM OF NEW PRACTICE

After describing the inherent effectiveness of automated techniques, we may now ask a question. Is the help from computer relevant and necessary for all types of content analytical research? This question should be asked because presently there is a craze among the analysts of certain countries of Western Europe and North America to consider the use of computer a *must* for any kind of content analytical work. Let us briefly put forward our arguments.

It is worthwhile to remember that no standard definition of content analysis makes it binding on the analyst to use computer. Rather, pioneers like Lasswell had expressed apprehension about mechanical preoccupation with automated methods which tended to divert the analytical process from theory and hypothesis (Lasswell *et al.*, 1952). It has often been emphasized that the analyst, before taking help from computer, should repeatedly ask himself 'what is my purpose?'

Once we accept the crucial role of theory and hypothesis in determining the measurement and inference process, it is necessary to point out two related issues. Firstly, there may be different stages in theory and hypothesis formation. There may be a stage of research when the investigator's particular theory and hypothesis are not fully worked out, and the relevant variables may not be entirely operationally defined. In this type of investigation, a descriptive phase usually precedes the hypothesis-generating phase: the researcher begins a project with an area that he wishes to investigate, describes his data, and then sets down a series of possible hypotheses to be tested in future research.

Secondly, we would remember that in certain cases the state of social

theory is such that it is often difficult to know beforehand which variables are likely to have the most explanatory power in a specific situation. Although our social scientific theories may lead us to the concepts we like to study, we lack an adequate theory of knowledge to direct us in finding the alternative signs that express a particular concept. What we have been trying to drive at is that in the field of content analysis there may be some areas where an exploratory study should preferably be conducted before applying to the text, preconceived standardised categories and procedures for collection of data with the help of computer.

AN ILLUSTRATION

In order to give example of an area where it is difficult to get help from computer at the state of collection of data from the text, we shall refer to a study (Dasgupta, 1974) in which, a method for systematic and quantitative analysis of views on societal change in India since the middle of the 18th century was developed. For this purpose an exploratory study was conducted by collecting and analysing views of a group of 37 selected authors. As our *a priori* information did not allow us to construct hypothesis on the variables of societal change in the given context, we deliberately kept open the scope of societal change; it implied that we exposed ourselves to an investigation of all possible alternatives of recorded change relating to social behaviour in India during a particular period of time, as expressed by a group of selected authors.

Taking the limitations of the field for granted, we turned to provide a framework in order to obtain from the text five basic items of information: (i) *subject* of social behaviour in the perspective of change, (ii) *time* of occurrence of the behaviour, (iii) *social group* involved in the behaviour, (iv) *place* of occurrence of the behaviour, and, (v) *nature of data* in support of the preceding items. The framework was improvised to receive the entry of distinct propositions on change in a pattern of cause-effect relationship and dissolve them into a set of interconnected items of information. Thus, 214 distinct views on change were collected from the books of the selected authors and entered into the framework, without the help of computer. However, after the transcription of 214 cases, the collected series of views were classified and codified, and then the information in numerical codes were transferred to punch cards, making them thereby amenable to machine tabulations. Ultimately, a quantitative analysis was carried out after obtaining tables with the help

of computer regarding variations within, and correlations between, some of the basic items of information involving the distinct views of the selected authors.

Once it is found that the improvised method is not altogether unsatisfactory to analyse views on societal change in a somewhat systematic and quantitative way, a second-stage study has been envisaged where the method can be applied in a delimited field of intensive research. With the help of knowledge gained from the first-stage study, it is now suggested that the scope of societal change will not be kept unspecified and views will be marshalled on one of the themes on which elementary information has been gathered by the exploratory investigation.

While we are, thus, planning to apply content analytical technique to an intensive area of research, the most important problem we are facing today is not the question of use of computer but that of formulation of relevant categories which would link our data with theory; at the same time, these should serve as a basis for drawing inference. Here it is relevant to recollect the views of Lasswell who had strongly advocated that 'the man who wishes to use content analysis . . . should step himself' into the text. 'Before he begins to count, he should read it to detect characteristic mechanisms and format. From this knowledge he should' be ready 'to set up his categories' (Lasswell *et al.*, 1952).

OTHER CURRENT PROBLEMS

Besides the controversy raging between the use of deductive automated method and that of inductive empirical approach, another important problem which has worried thinkers on content analysis (Stone *et al.*, 1966) is that it is sometimes found that many analysts, often preoccupied with measurement, have felt that they should stay at the level of facts and let the reader draw the conclusions. What is usually done by them is only rearrangement of analysed text and preliminary deductions in lieu of inference proper. This attitude must be refuted as content analysis should primarily be considered as a research tool to be used by the social scientist for making inferences. This is where content analysis differs from specialised cataloging activity connected with library information service.

One more issue which is now discussed among the content analysts is the relation of data archival activity with content analysis. It has been suggested that content analysis can help data archive whenever there is

difficulty in conducting questionnaires schedule to appraise the frame of reference of data (Dasgupta and Biswas, 1971). Recently, it is stressed that data archive should no longer be primarily considered as a servicing centre; data archival activity, like content analysis, should be conducted by a social scientist in connection with his own substantive research. Finally, it may be concluded that if we want content analysis to acquire its proper status among social sciences in India and elsewhere it should perform both of its functions simultaneously : measurement and inference. This implies that a social scientist, if he finds it relevant, should apply content analysis technique in his own substantive research problem in order to help himself in hypothesis generation, category formulation, collection and measurement of textual data, and inference process.

REFERENCES

- Dasgupta, A. K. 1974. "Views on change in Indian Society : A method for Analysis". In : *Paper presented to Workshop on Content Analysis (Pisa, Italy)*. Paris, UNESCO-International Social Science Council.
- Dasgupta, A. K. and S. K. Biswas, 1971. A Scheme for Regional Data Centre in India. *Working Paper of Sociological Research*, (Mimeographed). Indian Statistical Institute.
- Lasswell, H. D., N. Leites, and associates. 1949. *Language of Politics*. New York: Stewart.
- Lasswell, H. D., D. Lerner, and I. D. S. Pool, 1952. *The Comparative Study of Symbols*. Stanford University Press.
- Mukherjee, R. and associates. 1971. *Data Inventory on Social Sciences—India: First Phase (1967-68)*. Calcutta : Statistical Publishing Society.
- Stone, P. J., D. C., Dunphy, M. S. Smith, and D. M. Ogilvie, 1966. *The Ground Inquirer—A Computer Approach to Content Analysis*. Cambridge: Mass, I. I. T. Press.