



3-1987

Review of Robert Weber, "Basic Content Analysis"

Klaus Krippendorff

University of Pennsylvania, kkrippendorff@asc.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/asc_papers



Part of the [Communication Commons](#)

Recommended Citation (OVERRIDE)

Krippendorff, K. (1985). Basic Content Analysis, by Robert P. Weber. Beverly Hills, CA: Sage Publications, *Journal of the American Statistical Association* 82, 354-355. doi: 10.1080/01621459.1987.10478437

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/asc_papers/540
For more information, please contact repository@pobox.upenn.edu.

Review of Robert Weber, "Basic Content Analysis"

Disciplines

Communication | Social and Behavioral Sciences

Basic Content Analysis.

Robert Philip Weber. Beverly Hills, CA: Sage Publications, 1985. 95 pp. \$6.00 (paperback).

This is a short and inexpensive book intended to introduce both novices and somewhat more experienced researchers to increasingly available computer aids for content analysis. The role of content analysis in social inquiries needs no justification here. The growing availability of computers that could relieve the traditional content analyst of some of the tediousness of its procedures makes Weber's aim timely.

The first chapter is an introduction to content analysis as a method, its definition, some problems, and examples drawn from computer-aided studies.

The second chapter discusses coding or classification schemes. Although Weber gives an eight-step procedure for defining recording units and categories for use by human coders, the bulk of the chapter concerns problems of constructing dictionaries for computer applications. Since the publication of the General Inquirer system (Stone, Dunphy, Smith, and Ogilvie 1966), dictionaries serve to classify and assign analytically relevant tags to the words of a text. The *Harvard Psychosociological Dictionary*, originally introduced by Stone et al. (1966), and the *Lasswell Value Dictionary*, known since a study by Namenwirth (1969), dominate this chapter and much of the remainder of the book. Since the use of computers effectively eliminates problems of unreliability, which often threaten content analyses using human coders, Weber is concerned here with the validity of dictionaries largely resulting from the ambiguity of word meanings.

In the third chapter, Weber mentions sampling problems and possibilities of data entry by optical character readers, but he devotes the remainder of this largest chapter to a discussion of the usefulness of various computational aids: Key-Word-In-Context (KWIC) lists, word frequency counts, text retrieval by categories (tags), word counts by categories, factor analysis applications, and a structural model example. The assumptions underlying these techniques are taken for granted. This chapter, its illustrations, and the appendix correspond almost verbatim to Weber (1984) (except that the frequency tables purporting to characterize the same data are slightly different).

The fourth and concluding chapter cautions the researcher against interpreting the results carelessly, particularly by invoking problems of the semantic validity of categories and dictionaries, and shows how some researchers have tried to overcome shortcomings in procedure, as in Kelly and Stone's (1975) word disambiguation procedures, for example. The last pages of this chapter present an artificial intelligence approach to cognitive psychology as an example of what content analysis might look like in the far future.

The book is "basic" in that it does not presuppose the reader to have knowledge of computers, of linguistics, or of advanced statistical concepts. Moreover, the methodological issues of reliability and validity and a little bit of sampling are discussed in their most elementary form. The book is, therefore, easy to read. But this advantage also leads to a paradox. Those novices who feel invited to undertake a simple content analysis-for a small project or a term paper, for example-will find little practical guidance in that the techniques discussed in the book require accessibility of far-from basic computer software packages and a level of computational and analytical skills not widely shared among students and researchers in the social sciences. The book develops little appreciation for the amount of work that goes into the development of a suitable dictionary, for example. The appendix gives the sources for three word analysis systems, a tagging and retrieval system (TEXTPACK), a concordance program (OCP), and the KWIC system and it directs interested readers to the journal *Computers and the Humanities* for further inspiration. These may indeed contain the starters to get more seriously involved, but they are far from "basic."

Against the background of literature on content analysis, Weber's book can be regarded as a specialized update of material covered in other introductions. Berelson's (1952) original introduction, written in 1948, had little to say about computers. Lasswell, Lerner, and Pool (1952, p.63) expressed great hopes. Fifteen years later, at the Annenberg Conference, about half of the contributions, edited by Gerbner, Holsti, Krippendorff, Paisley, and Stone (1969), already concerned problems and applications of computers in various phases of content analysis. Holsti (1969) devoted the last chapter (or one-fifth of his book) to computer applications, reviewing word counting programs, KWIC lists, and programs with various kinds of dictionaries. He also discussed problems of homographs (words spelled alike but different in meaning), of data entry including machine-readable forms, and so forth, much as Weber does here. A chapter in Krippendorff's (1980) introduction also reviewed similar computational aids to content analysis, dictionary and thesaurus approaches, and a then-promising development in artificial intelligence, cognitive psychology, and computational linguistics. Weber's book essentially updates these chapters and expands on Weber (1984). Thus it is noteworthy that less than a third of its references are more recent than 1980. The artificial intelligence approach to cognition, reviewed in its infancy by Krippendorff and similarly heralded by Weber as a future form of content analysis, has, despite considerable progress, not yet found practical applications. The journal *Computers and the Humanities*, which Weber suggests his readers consult, already was one of five recommended in Holsti's introduction 16 years earlier. Serious readers might have liked to know what happened to the wide range of computational approaches and dictionaries presented in the Gerbner et al. (1969) volume. For examples-Did they not survive the evolution? Or why are the two Weber uses in his own work preferable to those mentioned in other literature? Perhaps Weber was well advised not to work into his book recent literature on theoretical and computational linguistics and on automatic text analysis, including translation. But progress is noticeable in these areas. Especially after the introduction of word processors, rather simple applications are spreading rapidly [see Conrad and Reinhartz (1984) for many and more recent examples].

Weber's primary justification for computer uses in content analysis is their unquestionable reliability. Readers who may ask themselves why this path has not generally been taken, however, find no good answers here. What the book fails to make clear is that this reliability is largely traded for the meanings human coders of a text can recognize easily (albeit sometimes unreliably), whereas computers generally do not. It is only in the last few pages that readers learn how much computers must know about language, psychological reasoning, interpersonal relations, and the discourse in question, to name but a few, before they can exhibit a reasonable level of "understanding" natural language. The kind of "understanding" that Weber uses in his own work is the knowledge of where a word begins and where it ends and in which predefined category it belongs. The latter decision is made sometimes after mechanically chopping off prefixes, suffixes, and grammatical endings and sometimes by looking for clue words in the environment of its sentence. The omission from analysis of syntax and the semantics of a word's linguistic context provides for a poor understanding of a text, and the reliability that such an analysis yields may be literally meaningless. What such a word-in-category analysis produces is something quite different from what human interpreters can recognize in text.

In the old debate over emic (indigenous) versus etic (investigator imposed) content categories, which is also echoed in the distinction between a posteriori (derived from text itself) and a priori (text independent) categories of analysis, Weber, citing comparability of results, sides with the latter. This preference is appropriately justified by large-scale predictions, such as those for particular value concerns. Consequently, the book ignores work aimed at finding out how text sources conceptualize their world and, in their own terms, how readers or particular classes of readers might interpret a text. This is the aim of much work in literature, ethnomethodology, and hermeneutics, among others, and provides an important source of understanding differences in the symbolic construction of social reality. Words in analytical categories that might not be the ones employed by their source may give only a very shallow indication of what that source says, is concerned with, or does.

In view of this limited form of understanding text, the book's title may be seriously misleading. Content analysis generally entails more. In his introduction, Weber agrees with other writers who make the drawing of valid inferences from text along the paths of its meanings and symbolic qualities a definitional requirement of content analysis. But then, probably because of the obvious difficulty of making computers understand natural language, he withdraws behind examples of what computers do more easily and that is treating words as units of analysis.

The work presented in Weber's book is not representative of what content analysts generally do. The reader should not confuse the aim of content analysis-probing into the symbolic worlds of individuals and groups-with what this book illustrates well-cutting a large text into separate words and putting these into fewer, theoretically justified, and ultimately predictive categories. "Computer word analysis," perhaps with the addition of "as aids to content analysis," might describe the gist of the book more appropriately.

KLAUS KRIPPENDORFF
University of Pennsylvania

REFERENCES

- Berelson, B. (1952), *Content Analysis in Communication Research*, New York: Free Press.
- Conrad, P., and Reinhartz, S. (eds.) (1984), "Computers and Qualitative Data," special issues of *Qualitative Sociology*, 7, Nos. 1 and 2.
- Gerbner, G., Holsti, O. R., Krippendorff, K., Paisley, W., and Stone, P. J. (eds.) (1969), *The Analysis of Communication Content: Development in Scientific Theories and Computer Techniques*, New York: John Wiley.
- Holsti, O. R. (1969), *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison-Wesley.
- Kelly, E. F., and Stone, P. J. (1975), *Computer Recognition of English Word Senses*, Amsterdam: North-Holland.
- Krippendorff, K. (1980), *Content Analysis: An Introduction to Its Methodology*, Beverly Hills, CA: Sage Publications.
- Lasswell, H. D., Lerner, D., and Pool, I. D. S. (1952), *The Comparative Study of Symbols*, Stanford, CA: Stanford University Press.
- Namenwirth, J. Z. (1969), "Some Long - and Short-Term Trends in One American Political Value: A Computer Analysis of Concern With Wealth in 62 Party Platforms," in *The Analysis of Communication Content: Development in Scientific Theories and Computer Techniques*, eds. G. Gerbner, O. R. Holsti, K. Krippendorff, W. Paisley, and P. J. Stone, New York: John Wiley, pp. 223- 241.
- Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (eds.) (1966), *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge, MA: MIT Press.
- Weber, R. P. (1984), "Computer-Aided Content Analysis: A Short Primer," *Qualitative Sociology*, 7, 126-147.