

Second Edition



Content Analysis

An Introduction to Its Methodology

Klaus Krippendorff

Second Edition

Content Analysis

An Introduction to Its Methodology

Klaus Krippendorff
University of Pennsylvania

 **SAGE Publications**
International Educational and Professional Publisher
Thousand Oaks ■ London ■ New Delhi

Copyright © 2004 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

For information:



Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

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

Sage Publications India Pvt. Ltd.
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
Krippendorff, Klaus.

Content analysis : an introduction to its methodology / Klaus Krippendorff.— 2nd ed.
p. cm.

Includes bibliographical references and index.
ISBN 0-7619-1544-3 — ISBN 0-7619-1545-1 (pbk.)

1. Content analysis (Communication) I. Title.
P93.K74 2004
401'.41—dc21

2003014200

Printed on acid-free paper.

03 04 05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Acquiring Editor: Margaret H. Seawell
Editorial Assistant: Jill Meyers
Production Editor: Claudia A. Hoffman
Copy Editor: Judy Selhorst
Typesetter: C&M Digitals (P) Ltd.
Indexer: Molly Hall
Cover Design: Concept by Klaus Krippendorff
Execution by: Michelle Lee Kenny

Contents

Preface to the Second Edition	xiii
Acknowledgments	xv
Introduction	xvii
Part I. Conceptualizing Content Analysis	
1. History	3
1.1 Some Precursors	3
1.2 Quantitative Newspaper Analysis	5
1.3 Early Content Analysis	6
1.4 Propaganda Analysis	8
1.5 Content Analysis Generalized	11
1.6 Computer Text Analysis	12
1.7 Qualitative Approaches	15
2. Conceptual Foundation	18
2.1 Definition	18
2.2 Epistemological Elaborations	21
2.3 Examples	26
2.4 Framework	29
2.4.1 Texts	30
2.4.2 Research Questions	31
2.4.3 Context	33
2.4.4 Analytical Constructs	34
2.4.5 Inferences	36
2.4.6 Validating Evidence	39
2.5 Contrasts and Comparisons	40

3. Uses and Inferences	44
3.1 Traditional Overviews	44
3.2 Extrapolations	47
3.2.1 Trends	49
3.2.2 Patterns	50
3.2.3 Differences	51
3.3 Standards	54
3.3.1 Identifications	54
3.3.2 Evaluations	55
3.3.3 Judgments	57
3.4 Indices and Symptoms	58
3.5 Linguistic Re-Presentations	62
3.6 Conversations	66
3.7 Institutional Processes	68
3.8 Areas of Likely Success	74

Part II. Components of Content Analysis

4. The Logic of Content Analysis Designs	81
4.1 Content Analysis Designs	81
4.1.1 Components	83
4.1.2 Quantitative and Qualitative Content Analysis	87
4.2 Designs Preparatory to Content Analysis	89
4.2.1 Operationalizing Knowledge	90
4.2.2 Testing Analytical Constructs as Hypotheses	91
4.2.3 Developing a Discriminant Function	91
4.3 Designs Exceeding Content Analysis	93
4.3.1 Comparing Similar Phenomena Inferred From Different Bodies of Texts	93
4.3.2 Testing Relationships Among Phenomena Inferred From One Body of Texts	94
4.3.3 Testing Hypotheses Concerning How Content Analysis Results Relate to Other Variables	95

5. Unitizing	97
5.1 Units	97
5.2 Types of Units	98
5.2.1 Sampling Units	98
5.2.2 Recording/Coding Units	99
5.2.3 Context Units	101
5.3 Ways of Defining Units	103
5.3.1 Physical Distinctions	103
5.3.2 Syntactical Distinctions	104
5.3.3 Categorial Distinctions	105
5.3.4 Propositional Distinctions	106
5.3.5 Thematic Distinctions	107
5.4 Productivity, Efficiency, and Reliability	109
6. Sampling	111
6.1 Sampling in Theory	111
6.2 Sampling Techniques Applicable to Texts	113
6.2.1 Random Sampling	114
6.2.2 Systematic Sampling	115
6.2.3 Stratified Sampling	115
6.2.4 Varying Probability Sampling	116
6.2.5 Cluster Sampling	116
6.2.6 Snowball Sampling	117
6.2.7 Relevance Sampling	118
6.2.8 Census	120
6.2.9 Convenience Sampling	120
6.3 Sample Size	121
6.3.1 Statistical Sampling Theory	121
6.3.2 Sampling Experiments	122
6.3.3 The Split-Half Technique	124
7. Recording/Coding	125
7.1 The Function of Recording and Coding	125

7.2	Coder Qualifications	127
7.2.1	Cognitive Abilities	127
7.2.2	Background	128
7.2.3	Frequency	128
7.3	Coder Training	129
7.4	Approaches to Defining the Semantics of Data	132
7.4.1	Verbal Designations	132
7.4.2	Extensional Lists	133
7.4.3	Decision Schemes	135
7.4.4	Magnitudes and Scales	136
7.4.5	Simulation of Hypothesis Testing	137
7.4.6	Simulation of Interviewing	139
7.4.7	Constructs for Closure	141
7.5	Records	143
7.5.1	Administrative Information	144
7.5.2	Information on the Organization of Records	145
7.5.3	Substantive Information About the Phenomena of Interest	146
8.	Data Languages	150
8.1	The Place of Data Languages in Analysis	150
8.2	Definitions	153
8.3	Variables	155
8.4	Nominal Variables	161
8.5	Orderings	161
8.5.1	Chains	162
8.5.2	Recursions	163
8.5.3	Cubes	163
8.5.4	Trees	164
8.6	Metrics	165
8.6.1	Ordinal Metrics	166
8.6.2	Interval Metrics	168
8.6.3	Ratio Metrics	169
8.7	Mathematical Operations	169

9. Analytical Constructs	171
9.1 The Role of Analytical Constructs	171
9.2 Sources of Certainty	173
9.2.1 Previous Successes and Failures	173
9.2.2 Expert Knowledge and Experience	175
9.2.3 Established Theories	176
9.2.4 Embodied Practices	179
9.3 Types of Constructs	179
9.3.1 Extrapolations	180
9.3.2 Applications of Standards	180
9.3.3 Indices and Symptoms	180
9.3.4 Re-Presentations	182
9.3.5 Conversations/Interactions	183
9.3.6 Institutional Processes	183
9.4 Sources of Uncertainty	185
9.4.1 Variance of the Target	185
9.4.2 Confidence Levels	186
9.4.3 Appropriateness of the Construct	186

Part III. Analytical Paths and Evaluative Techniques

10. Analytical/Representational Techniques	191
10.1 Tabulations	192
10.2 Cross-Tabulations, Associations, and Correlations	194
10.3 Multivariate Techniques	197
10.4 Factor Analysis and Multidimensional Scaling	200
10.5 Images, Portrayals, Semantic Nodes, and Profiles	202
10.6 Contingencies and Contingency Analysis	205
10.7 Clustering	208
11. Reliability	211
11.1 Why Reliability?	211
11.2 Reliability Designs	214

11.2.1	Types of Reliability	214
11.2.2	Conditions for Generating Reliability Data	216
11.2.3	Reliability Data	219
11.3	α-Agreement for Coding	221
11.3.1	Two Observers, Binary Data	223
11.3.2	Two Observers, Many Nominal Categories	227
11.3.3	Many Observers, Many Nominal Categories, Missing Values	230
11.3.4	Data With Different Metrics	232
11.4	Statistical Properties of α	236
11.4.1	Insufficient Variation	236
11.4.2	Statistical Significance	237
11.4.3	Sampling Considerations	238
11.4.4	Standards for Data Reliability	241
11.5	Other Coefficients and Correspondences	244
11.6	α-Agreement for Unitizing	251
12.	Computer Aids	257
12.1	What Computers Do	257
12.2	How Computers Can Aid Content Analyses	258
12.3	Accounts of Character Strings	262
12.4	Text Searches	272
12.5	Computational Content Analyses	281
12.5.1	Coding/Dictionary Approaches	283
12.5.2	Statistical Association Approaches	289
12.5.3	Semantic Network Approaches	292
12.5.4	Memetic Approaches	298
12.6	Interactive-Hermeneutic Explorations	303
12.7	Frontiers	307
12.7.1	Intelligent Browsers	307
12.7.2	Common Platforms	308
12.7.3	Computational Theories of Meaning	309

12.7.4	Utilization of Intertextualities	309
12.7.5	Natural Interfaces	310
13.	Validity	313
13.1	Validity Defined	313
13.2	A Typology for Validating Evidence	318
13.2.1	Sampling Validity	321
13.2.2	Semantic Validity	323
13.2.3	Structural Validity	330
13.2.4	Functional Validity	332
13.2.5	Correlative Validity	333
13.2.6	Predictive Validity	336
14.	A Practical Guide	339
14.1	Designing an Analysis	340
14.1.1	Text-Driven Analyses	341
14.1.2	Problem-Driven Analyses	342
14.1.3	Method-Driven Analyses	355
14.2	Writing a Research Proposal	357
14.2.1	Rhetorical Function	358
14.2.2	Contractual Function	359
14.2.3	Outline for a Research Proposal	359
14.3	Applying the Research Design	361
14.4	Narrating the Results	362
14.4.1	Outline for a Research Report	363
	References	365
	Index	391
	About the Author	413

Preface to the Second Edition

Content analysis is potentially one of the most important research techniques in the social sciences. The content analyst views data as representations not of physical events but of texts, images, and expressions that are created to be seen, read, interpreted, and acted on for their meanings, and must therefore be analyzed with such uses in mind. Analyzing texts in the contexts of their uses distinguishes content analysis from other methods of inquiry.

Methods in the natural sciences are not concerned with meanings, contents, intentions, and references. These scientists hardly reflect on their own conceptions of nature, excluding their conceptions from their object of study by dismissing them as subjective in contrast to what can be determined through detached observation and objective measurement. Where social researchers adopt natural scientific methods of inquiry, the epistemology that is inscribed in such methods prevents them from addressing what matters most in everyday social life: human communication, how people coordinate their lives, the commitments they make to each other and to the conceptions of society they aspire to, what they know, and why they act. Certainly, content analysis is not the only research method that takes meanings seriously, but it is a method that is both powerful and unobtrusive. It makes sense of what is mediated between people—textual matter, symbols, messages, information, mass-media content, and technology-supported social interactions—without perturbing or affecting those who handle that textual matter.

In the first edition of *Content Analysis*, published in 1980, I suggested that content analysis was at a crossroads. Content analysts at that time had a choice: They could continue their shallow counting game, motivated by a journalistic fascination with numbers and a narrow conception of science in which quantitative measurement provides the only evidence that counts (Lasswell, 1949/1965b), or they could refocus content analysis methods on social phenomena that are both generated by and constituted in texts and images and, hence, need to be understood through their written and pictorial constituents. Although the logic and methods that I presented in the first edition of *Content Analysis* have survived their challenges, the textual fabric of contemporary society has undergone radical transformations, due in no small part to the ongoing information revolution. The increasingly widespread availability of electronic, and hence computer-readable,

texts concerning virtually everything that matters to society and its members has moved content analysis, particularly computer-aided text analysis, into the center of how society examines itself.

In the 1980s, content analysis was a research method that had entered the psychological and social sciences, but was used mainly in journalism and communication research. At that time, the amount of human effort required to collect, transcribe, and code textual data made content analysis a time-consuming and labor-intensive effort. Today, content analysis has become an efficient alternative to public opinion research, a method of tracking markets, political leanings, and emerging ideas; it is used as a way to settle legal disputes and as an approach to the exploration of individual human minds—not to dwell on the many improvements that content analysts have made in traditional content analytic inquiries of the mass media. Despite remarkable progress, content analysts can hardly claim to have met the challenges of this new era. The imagined analytical potential is far ahead of what can be done today, fueling the work of many developers of new analytic tools.

Although the outline of this new edition remains essentially unchanged from that of the first, this volume clarifies numerous methodological issues in content analysis and responds to the technique's latest challenges. Accordingly, I have substantially rewritten all chapters, addressing developments that have taken place since 1980, especially Chapter 12, on computer-aided text analysis, and Chapter 14, a practical guide, which incorporates my experiences in teaching and consulting on academic and commercial research projects. I have also substantially revised my earlier discussions of the epistemology, logic, and methods of content analysis.

I thank my students at the University of Pennsylvania's Annenberg School for Communication for their open minds and my colleagues for presenting me with the challenging problems of their content analyses. I would also like to thank numerous readers of the first edition—both students and practicing content analysts—for sharing their comments and criticisms, and Sage Publications for giving me more space for this edition.

The first edition of *Content Analysis* has been translated into Italian, Japanese, Spanish, and Hungarian, and during the 23 years since its publication, it has reached an enormous audience. It has been widely adopted as a text in social science, humanities, and business curricula. It has served researchers as a guide to the design and execution of large and small content analyses, and it has provided a standard for justifying as well as critically evaluating content analysis findings. When I travel to national and international conferences, I continue to be amazed and pleased to meet researchers from all over the world who tell me how studying this text has helped them in their current inquiries. This new edition is written for the same wide audience of practicing researchers, social scientists, and students.

—Klaus Krippendorff
Gregory Bateson Term Professor for Cybernetics, Language, and Culture
The Annenberg School for Communication
University of Pennsylvania

Acknowledgments

This book is dedicated to scholars—beginning and established—who, while critical of any methodology as they should be, nevertheless are willing to add another perspective to their reading of especially voluminous textual matter.

I would like to thank my students at the University of Pennsylvania's Annenberg School for Communication for teaching me over the years what is important in a textbook on content analysis and for offering invaluable feedback on a draft of this new edition. I am grateful also for thoughtful reviews of the manuscript by William Benoit, Wayne Danielson, Gavan Duffy, William Evans, Kenneth Janda, and Mark West. In particular, I thank Kenneth Janda for his most thorough critique and William Benoit for not only making detailed recommendations but also giving the manuscript a test run in his course.

Introduction

The term *content analysis* is about 60 years old. *Webster's Dictionary of the English Language* included the term in its 1961 edition, defining it as “analysis of the manifest and latent content of a body of communicated material (as a book or film) through classification, tabulation, and evaluation of its key symbols and themes in order to ascertain its meaning and probable effect.” The intellectual roots of content analysis, however, can be traced far back in human history, to the beginning of the conscious use of symbols and voice, especially writing. This conscious use, which replaced the magical use of language, has been shaped by the ancient disciplines of philosophy, rhetoric, and cryptography. It has also spawned religious inquisitions and political censorship on the part of ruling establishments. Today, symbolic phenomena are institutionalized in art, literature, education, and the mass media, including the Internet. Theoretical and analytical concerns are found in such academic disciplines as anthropology, linguistics, social psychology, sociology of knowledge, and the comparatively younger field of communication studies. Many practical pursuits have grown from these fields: psychotherapy, advertising, politics, the arts, and so on. Virtually all disciplines within the whole spectrum of the humanities and the social sciences, including those that seek to improve the political and social conditions of life, are concerned with the functions and effects of symbols, meanings, and messages. In recent years, the emergence of the information society has moved the minutiae of communication—texts, contexts, images, interfaces, and, above all, information—into the very center of researchers’ attempts at self-understanding.

However ancient the roots of analyzing symbolic and textual matter might be, today’s content analysis is significantly different, in aim and in method, from that of the past. Contemporary content analysis has three distinguishing characteristics.

First, content analysis is an *empirically grounded method*, exploratory in process, and predictive or inferential in intent. Many of our current concepts relating to language are of Greek origin; for example, the words *sign*, *significance*, *symbol*, and *logic* all have Greek roots. However, the ancient Greeks’ interest in language was largely prescriptive and classificatory, not empirical. Aristotelian logic set the standards for clear expression, and much of rhetorical theory was directed toward a normative conception of persuasive argumentation. Science that explores rather than declares is a relatively recent accomplishment. Only a century ago, George Boole and his contemporaries believed that the brain works according to (Boolean) logic and that human conduct is entirely rational. However, computers built on this logic turned out to be rather disappointing

thinking machines. Empirical research in psychology is replacing Aristotelian categories in favor of a “psycho-logic.” And we no longer measure human communication against the ideal of transmitting information. Instead, we inquire into what happens to the relationships between people who converse with one another.

With new conceptualizations and an empirical orientation, contemporary content analysts join other researchers in seeking valid knowledge or practical support for actions and critique. However, unlike researchers who employ other empirical techniques, content analysts examine data, printed matter, images, or sounds—texts—in order to understand what they mean to people, what they enable or prevent, and what the information conveyed by them does. These are questions for which natural scientists have no answers and for which their methods are generally insensitive.

Second, contemporary content analysis *transcends traditional notions of symbols, contents, and intents*. This may be seen in the evolution of the concept of communication, in how the development of media technologies has shaped our attention to communication, and in the role of culture in assigning significance to what is being analyzed. I would argue that in recent years our awareness of communication has undergone four conceptual revolutions, as described below, and probably is in the midst of a fifth:

- The idea of *messages*: the early awareness not only that verbal discourse is movable when written, but that writing has predictable effects. This awareness emerged in ancient Greece when messengers were used as the carriers of significance, history became documented, laws of the land were laid down in writing, and written instructions built organizational structures, directed events, and influenced (and possibly deceived) their receivers or the public. The concept of a message was a precursor of the rhetorical exploration of language. Tropes, syllogisms, and meanings came to be thought of as inherent qualities of speeches, letters, or documents. But a message is the metaphorical container of all these, a “container of content,” a vehicle for shipping meanings from one place to another—for example, when we now leave a message for someone on an answering machine or say that a message was meaningful (full of meanings) or meaningless (void of meanings).
- The idea of *channels*: the awareness of the constraints that every medium imposes on human communication. This awareness came with the increased reliance on different media of communication and served to explain their limitations: The alphabet limits what one can say in writing; the telephone confines communication to sound; and a television station can air no more than what is transmittable without interference from other stations, appealing to large audiences, and deemed profitable by its sponsors. The channel metaphor conjures images of canals and pipes with restricted capacities for shipping messages (with their contents) of certain forms and volumes.

- The idea of *communication*: the awareness of the relational space between senders and receivers, of the processes through which interpersonal relations are negotiated, social structures are constituted, and members of large populations come to know about each other. This awareness developed as an offshoot of the growth in mass media. By producing and disseminating identical messages—news and entertainment—to everyone, the mass media promised to be an agent of sharing, of building community relationships, of democratization, ideally, worldwide. Modeling themselves on the idea of mass production, the mass media also made us aware of where this one-way model failed: in interpersonal conversation, point-to-point telephone communication, public debate, and dialogue. In U.S. culture, mass-media technology has become synonymous with progress, and communication is understood as the cure for most social problems—for example, we often blame lack of communication or miscommunication when interpersonal as well as national conflicts arise.
- The idea of *systems*: the awareness of global, dynamic, and technologically supported interdependencies. This idea emerged with the growth of communication networks—telephone nets, wire services, mass-media systems, and most recently the Internet—transforming commerce, politics, and interpersonal relationships, creating networks whose properties have so far defied attempts to theorize them adequately. Unlike the one-way mass media, systems are marked by the interactivity and simultaneity of parallel communication on a massive scale and with the potential of nearly universal participation.
- The idea of *computation*: the awareness of the algorithmic nature of certain routine cognitive and social processes and their implementation in increasingly powerful computers. The processing of digital data in place of cognitive and social practices, along with the ability to reproduce these data in visual and textual forms for reading, rearticulating, and disseminating by and to ideally everyone, is encouraging an entirely new literacy that undercuts traditional organizational structures, including national boundaries. The fluidity and enormous complexity that computation has introduced into almost all spheres of life amplify the possibilities for scientific exploration as well as present unprecedented challenges for collective understanding.

This rather sketchy history of communication suggests that researchers who are concerned with texts can no longer focus only on symbols or representations, nor can they limit themselves to questions about “who says what, through which channels, to whom, and with which effects” (Lasswell, 1960). The popular and simplistic notion of “content” has outlived its explanatory capabilities as well: *content*, the *what* of a communication, an entity that authors think they *enter* into messages and *ship* to remote receivers, who *remove* it for *what it is* and henceforth

share it among others. This bizarre notion leads to authors as authorities of what they put into messages and to the conception of *content analysts* as experts who provide objective accounts of what messages were *intended to convey* or *actually contain*.

The virtuality of electronic media encourages short-lived access to messages that, without knowledge of their human authors, calls for a new technological basis for trust. It coordinates the lives of many people, overcoming old distinctions among channels of communication, obviating physical distances, and pushing capacities of the human participants to their limits. This erodes the validity of traditional communication theories, all the while enabling computer systems to thrive in this new environment. It is these computer systems that simulate and coordinate parts of the very social processes that researchers wish to understand. This is a radically changing world in which texts play distinctly new roles. Newspaper accounts, public opinion polls, corporate reports, files in government agencies, credit information, bank transactions, and, above all, huge textual data archives—all are now linked into networks that can be analyzed from numerous positions. In effect, the social systems that we conceived of as explaining society are now holographically retreating into our computers. This development calls for a redefinition of content analysis, one that aligns content—the target of the research—with how contemporary society operates and understands itself through its texts.

With the container metaphor rendered useless, perhaps the term *content analysis* no longer fits the realities of contemporary society. For better or for worse, I continue to use the term in this book, but I also plead with readers to oppose unflinchingly the naive and misleading entailments of the pervasive container metaphor.

Third, *contemporary content analysis has been forced to develop a methodology of its own*, one that enables researchers to plan, execute, communicate, reproduce, and critically evaluate their analyses whatever the particular results. Content analysts have had to develop such a methodology for three reasons:

- *Content analysts now face larger contexts.* The shift in interest from small collections of printed messages to systems and then to electronic texts and images circulating in the environment of content analysts is tied less to the nature of textual data than to the increasingly complex worlds that produce and are sustained by these data. This shift calls for theories and conceptions that earlier content analysts did not need. Although content analysts have frequently lamented the lack of general theories that could justify their work, progress in implementing more specific or micro-level theories is encouraging. This is especially true where content analysis has migrated through disciplines that were not previously concerned with textual data, such as the cognitive sciences and artificial intelligence.
- *Greater numbers of researchers need to collaborate in the pursuit of large-scale content analyses.* This observation is a correlate of the growing sample sizes of relevant texts, the analysis of which easily exceeds what individual

analysts can handle. It implies that content analysts must work together, in parallel, and as research teams. Teamwork, however, needs to be organized reliably. Both the social problem of coordinating researchers and the methodological problem of assuring replicability tend to be solved through the adoption of a language whose vocabulary enables researchers to clarify the analytical procedures they use, negotiate the individual responsibilities of the participants, assure agreement on the analytical categories, and evaluate the performance of team members.

- *The large volumes of electronically available data call for qualitatively different research techniques*, for computer aids. Such aids convert large bodies of electronic text into representations if not answers to research questions that content analysts need to understand. However, exactly what sophisticated text analysis software does—aside from promising to carry out the more labor-intensive clerical parts of processing textual data—is often difficult to retrace and inaccessible to the average content analyst. These computer aids participate in content analysis much as human analysts do. They become part of its methodology, with transparency being a major issue.

To be clear, *methodology* is not a value in itself. The purpose of methodology is to enable researchers to plan and examine critically the logic, composition, and protocols of research methods; to evaluate the performance of individual techniques; and to estimate the likelihood of particular research designs to contribute to knowledge. Every researcher must become proficient in defining the terms of an analysis and justifying the analytical steps taken to a skeptical friend or questioning colleague. Methodology provides a language for talking about the process of research, not about subject matter. In the history of scientific pursuits, the development of methodology has always been a major accomplishment. For example, for thousands of years humans preserved history by retelling or chanting stories, since the *Iliad* in writing, before the historian Leopold von Ranke, only a century ago, gave the “document” the methodological status it now has in the academic study of history. Similarly, scholars practiced “content analysis” well before Berelson and Lazarsfeld (1948) undertook the first codification of this method. Although many observers have argued that each content analysis is unique, possibly focusing largely on its subject matter, I would argue that all content analyses share a procedural logic and need to be justified through the use of socially acceptable criteria. These commonalities form the substance of this book.

I disagree with the frequent contention that content analysis is “nothing more than what everyone does when reading a newspaper, except on a larger scale.” Content analysis may have been that way, in its early, journalistic stage, and its methodology does not rule out such readings, but this narrow definition is no longer sufficient today. As newspaper readers, we are perfectly justified in applying our individual worldviews to texts and enacting our interest in what those texts mean to us; in fact, we cannot do otherwise. But as content analysis researchers, we must do our best to explicate what we are doing and describe

how we derive our judgments, so that others—especially our critics—can replicate our results.

This book, then, introduces readers to ways of analyzing meaningful matter, texts, images, and voices—that is, data whose physical manifestations are secondary to what they mean to particular populations of people. The chapters are grouped into three main parts. Part I, “Conceptualizing Content Analysis,” begins with a brief chapter on the history of content analysis. In Chapter 2, I develop a definition of content analysis that distinguishes this technique from other methods of inquiry, and in Chapter 3, I present a discussion of some of the ways in which content analysis has been applied. The chapters in Part II, “Components of Content Analysis,” outline the procedures used in content analyses, beginning with their procedural logic and moving naturally from unitizing to sampling, recording/coding, data languages, and analytical constructs. The chapters in Part III, “Analytical Paths and Evaluative Techniques,” trace several paths through content analysis protocols. In this part of the book, I discuss analytical constructs that enable researchers to draw inferences from data, the use of computers and computational techniques, and the two principal criteria used in evaluating content analyses: reliability and validity. In the final chapter, I provide a practical guide that summarizes the foregoing discussion from a practitioner’s perspective.

Readers who have never done a content analysis may want to begin by reading Chapter 1, on the history of content analysis, and Chapter 3, on the uses of this technique, to get a sense for whether or not it suits their research interests. If it does, they should familiarize themselves with the conceptual foundations of content analysis by reading Chapter 2. Beginners in content analysis are advised to start with a small pilot project, to get a feel for what is involved in conducting a larger study. Methodology without some practice is empty. The guidelines in Chapter 14, although written as a summary, could also serve as a start. In this chapter, readers will find many helpful references to pertinent chapters in this volume, which may answer emerging questions and place these answers within the context of larger methodological issues. Beginning researchers will soon realize that analyzing text is not a mechanical task, and neither is designing a content analysis. Both undertakings require creativity and competence.

Readers who have had some experience with coding will acquire a larger perspective on what they had been doing. As the table of contents suggests, coding is only a small part of content analysis—despite popular misconceptions. In fact, only Chapter 7 is devoted to issues of coding or recording, something researchers need do only when their data or texts are unwieldy. By coding/recording textual matter, one learns to appreciate both the conceptual problems involved in imposing analytical categories on ordinary readings of text and the ways in which competent researchers have managed to solve such problems. Designing a content analysis is something different, however. I recommend that readers who have had experience with coding expand on that experience by examining the chapters offered here about all the other components of content analysis, adding these to their conceptual frameworks. Such readers might well look into Chapter 12, on computer aids, to gain an alternative perspective on coding.

Readers who have already undertaken content analyses or similar text-based research will discover in this book alternative paths for such inquiries and a vocabulary that they can use in deliberating about what is involved in analyzing texts—not as observations of naturalistic phenomena, but as data whose significance stems from the meanings that others bring to their readings. Those who think they know what content analysis is are advised to start with Chapter 2, on the conceptual foundations of content analysis. This chapter discusses the ways that researchers talk about content and exposes readers to the larger perspective they will need in order to conceive a content analysis or critically evaluate the content analyses of others. As a condition for publication, scholarly journals increasingly demand some demonstration of why a content analysis should be taken seriously. In the past, content analysts relied heavily on conceptions of content as “contained” in messages, as discussed above, or “inherent” to texts. This settled the thorny issue of multiple text interpretations by fiat and consequently disabled explicitness about the researchers’ procedures. Several research traditions—such as interpretive research, discourse analysis, literary scholarship, and rhetoric—tend to be plagued by similar conceptions. Researchers from these traditions would greatly benefit from explicating their approaches, checking their results against the work of others, and evaluating the social consequences of their findings outside their own schools of thought—as I am suggesting.

For experts in content analysis, this book raises several epistemological questions that practitioners rarely ask, transforms them into methodological ones, and provides new solutions to practical problems.

Readers who must make decisions concerning whether or not to trust the findings of content analyses and other text-based research—for instance, judges in courts of law, practitioners in the fields of public relations and advertising, and reviewers of research submitted for funding or publication in scientific journals—will find the vocabulary of this book useful as they need to weigh the quality of findings and make informed recommendations for improvements. Such readers will find the discussions in Chapters 2, 11, and 13 (on conceptual foundations, reliability, and validity, respectively) especially applicable to their evaluative endeavors.

While this book may serve as a handbook for various practitioners, it grew out of my experiences in teaching courses and seminars in content analysis, and I conceive of it foremost as a textbook for advanced undergraduate and beginning graduate students. Teachers and their students may not want to work through all the chapters in their numerical order; for instance, those intending to use computers will find Chapter 12 more important than Chapter 7, on recording/coding, and may omit Chapter 11, on reliability issues. Students with specific projects in mind may pass over sections that may not be useful to their projects. However, readers should not rule out chapters as irrelevant before knowing the possibilities they offer.

Finally, for me, the book will have achieved its purpose if it helps to make the newly acquired wealth of textual data accessible to systematic analysis, if it improves the social significance of research in the humanities and the social sciences, and if it furthers the development of methods of inquiry into the realities that human communication constructs.