Fourth Edition

Content Analysis

An Introduction to Its Methodology

Klaus Krippendorff

Content Analysis

An Introduction to Its Methodology

Fourth Edition

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Content Analysis

An Introduction to Its Methodology

Fourth Edition

Klaus Krippendorff
Annenberg School for Communication
University of Pennsylvania

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

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

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Acquisitions Editor: Terri Accomazzo
Editorial Assistant: Erik Helton
Production Editor: Andrew Olson
Copy Editor: Michelle Ponce
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Ellen Howard
Indexer: Jeanne Busemeyer
Cover Designer: Anupama Krishnan
Marketing Manager: Liz Thornton

Copyright © 2019 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.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Krippendorff, Klaus, author.

Title: Content analysis : an introduction to its methodology / Klaus Krippendorff, University of Pennsylvania.

Description: Fourth Edition. | Los Angeles : SAGE, [2018] | Includes bibliographical references and index.

Identifiers: LCCN 2017050739 | ISBN 9781506395661 (Paperback : acid-free paper)

Subjects: LCSH: Content analysis (Communication)

Classification: LCC P93 .K74 2018 | DDC 302.2—dc23
LC record available at https://lccn.loc.gov/2017050739

This book is printed on acid-free paper.

18 19 20 21 22 10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

Preface to the Fourth Edition **xii**

Introduction 1

PART I • CONCEPTUALIZING CONTENT ANALYSIS

Chapter 1 • History 10

Chapter 2 • Conceptual Foundation 24

Chapter 3 • Uses and Inferences 51

PART II • COMPONENTS OF CONTENT ANALYSIS

Chapter 4 • The Logic of Content Analysis Designs 86

Chapter 5 • Unitizing 102

Chapter 6 • Sampling 115

Chapter 7 • Recording/Coding 128

Chapter 8 • Data Languages 157

Chapter 9 • Analytical Constructs 178

PART III • ANALYTICAL PATHS AND EVALUATIVE TECHNIQUES

Chapter 10 • Analytical/Representational Techniques 196

Chapter 11 • Computer Aids 215

Chapter 12 • Reliability 277

Chapter 13 • Validity 361

Chapter 14 • A Practical Guide 383

Glossary **407**

References **414**

Index **433**

About the Author **453**

DETAILED CONTENTS

Preface to the Fourth Edition xii

Introduction 1

PART I • CONCEPTUALIZING CONTENT ANALYSIS

Chapter 1 • History 10

1.1 Some Precursors 10
1.2 Quantitative Newspaper Analysis 11
1.3 Early Content Analysis 12
1.4 Propaganda Analysis 14
1.5 Content Analysis Generalized 17
1.6 Computer Text Analysis 18
1.7 Qualitative Approaches 21

Chapter 2 • Conceptual Foundation 24

2.1 Definition 24
2.2 Epistemological Elaborations 27
2.3 Examples 31
2.4 Framework 37
2.4.1 Texts 38
2.4.2 Research Questions 39
2.4.3 Context 40
2.4.4 Analytical Constructs 42
2.4.5 Inferences 43
2.4.6 Validating Evidence 45
2.5 Contrasts and Comparisons 46

Chapter 3 • Uses and Inferences 51

3.1 Traditional Overviews 51
3.2 Extrapolations 54
3.2.1 Trends 55
3.2.2 Patterns 56
3.2.3 Differences 58
3.3 Standards 61
3.3.1 Identifications 61
3.3.2 Evaluations 62
3.3.3 Judgments 64
3.4 Indices and Symptoms 65
3.5 Linguistic Re-Presentations 69

3.6 Conversations 73
3.7 Institutional Processes 74
3.8 Areas of Likely Success 80

PART II • COMPONENTS OF CONTENT ANALYSIS

Chapter 4 • The Logic of Content Analysis Designs 86

4.1 Content Analysis Designs 86
4.1.1 Components 87
4.1.2 Quantitative and Qualitative Content Analysis 91
4.2 Designs Preparatory to Content Analysis 94
4.2.1 Operationalizing Available Knowledge of the Context 94
4.2.2 Testing Analytical Constructs as Hypotheses 95
4.2.3 Developing Analytical Constructs by Trial and Error 96
4.3 Designs Going Beyond Content Analysis 98
4.3.1 Comparing Similar Phenomena Inferred From Different Bodies of Texts 99
4.3.2 Testing Relationships Among Phenomena Inferred From One Body of Texts 99
4.3.3 Testing Hypotheses Concerning How Content Analysis Results Relate to Other Variables 100

Chapter 5 • Unitizing 102

5.1 Units 102
5.2 Types of Units 103
5.2.1 Sampling Units 103
5.2.2 Recording/Coding Units 104
5.2.3 Context Units 105
5.3 Ways of Defining Units 107
5.3.1 Physical Distinctions 107
5.3.2 Syntactical Distinctions 108
5.3.3 Categorial Distinctions 109
5.3.4 Propositional Distinctions 110
5.3.5 Thematic Distinctions 111
5.4 Productivity, Efficiency, and Reliability 112

Chapter 6 • Sampling 115

6.1 Sampling in Theory 115
6.2 Sampling Techniques Applicable to Texts 117
6.2.1 Random Sampling 118
6.2.2 Systematic Sampling 118
6.2.3 Stratified Sampling 119
6.2.4 Varying Probability Sampling 119
6.2.5 Cluster Sampling 120
6.2.6 Snowball Sampling 121
6.2.7 Relevance Sampling 122
6.2.8 Census 123
6.2.9 Convenience Sampling 124

6.3 Sample Size 124
6.3.1 Statistical Sampling Theory 124
6.3.2 Sampling Experiments 125
6.3.3 The Split-Half Technique 127

Chapter 7 • Recording/Coding 128

7.1 The Function of Coding and Recording 128
7.2 Coder Qualifications 130
7.2.1 Cognitive Abilities 130
7.2.2 Background 130
7.2.3 Frequency 131
7.3 Coder Training 131
7.4 Crowdcoding 134
7.4.1 Methodological Advantages 135
7.4.2 Efforts to Assure Crowdcoding Quality 136
7.4.3 Uses of Reliability Tests 136
7.4.4 Limitations and Warnings 137
7.5 Approaches to Defining the Semantics of Data 138
7.5.1 Verbal Designations 139
7.5.2 Extensional Lists 140
7.5.3 Examples 141
7.5.4 Decision Schemes 142
7.5.5 Magnitudes and Scales 143
7.5.6 Simulation of Hypothesis Testing 145
7.5.7 Simulation of Interviewing 146
7.5.8 Constructs for Closure 148
7.6 Records 149
7.6.1 Administrative Information 151
7.6.2 Information on the Organization of Records 151
7.6.3 Substantive Information About the Phenomena of Interest 153

Chapter 8 • Data Languages 157

8.1 The Place of Data Languages in Analytical Efforts 157
8.2 Definitions 159
8.3 Variables 161
8.4 Nominal Variables 167
8.5 Ordered Variable 167
8.5.1 Chains 168
8.5.2 Recursions 169
8.5.3 Cubes 169
8.5.4 Trees 171
8.5.5 Multi-Valued Sets 171
8.6 Metrics 172
8.6.1 Ordinal Metrics 173
8.6.2 Interval Metrics 175
8.6.3 Ratio Metrics 176
8.7 Mathematical Operations 176

Chapter 9 • Analytical Constructs **178**

9.1 The Role of Analytical Constructs 178

9.2 Sources of Certainty 180

9.2.1 Previous Successes and Failures 180

9.2.2 Expert Knowledge and Experience 181

9.2.3 Established Theories 183

9.2.4 Embodied Practices 185

9.3 Types of Constructs 186

9.3.1 Extrapolations 186

9.3.2 Applications of Standards 186

9.3.3 Indices and Symptoms 186

9.3.4 Re-Presentations 188

9.3.5 Conversations/Interactions 189

9.3.6 Institutional Processes 189

9.4 Sources of Uncertainty 191

9.4.1 Variance of the Target Domain 191

9.4.2 Confidence Levels 192

9.4.3 Stability of the Analytical Constructs Model 192

PART III • ANALYTICAL PATHS AND EVALUATIVE TECHNIQUES

Chapter 10 • Analytical/Representational Techniques **196**

10.1 Counts 197

10.2 Cross-Tabulations, Associations, and Correlations 200

10.3 Multivariate Techniques 203

10.4 Factor Analysis and Multidimensional Scaling 205

10.5 Images, Portrayals, Semantic Nodes, and Profiles 207

10.6 Contingencies and Contingency Analysis 210

10.7 Clustering 212

Chapter 11 • Computer Aids **215**

11.1 What Computers Do 215

11.2 How Computers Can Aid Content Analyses 216

11.3 Text Analyses 220

11.3.1 Accounts of Character Strings 220

11.3.2 Text Mining 227

11.3.3 Mining Textual Evidence for Theories 236

11.3.4 Analysis of Networked Texts 239

11.4 Computational Content Analyses 246

11.4.1 Coding/Dictionary Approaches 248

11.4.2 Machine Learning of Discriminant Functions 254

11.4.3 Statistical Association Approaches 256

11.4.4 Semantic Network Approaches 258

11.4.5 Memetic Approaches 264

11.5 Qualitative Data Analysis Support 268

11.6 Frontiers 272
11.6.1 Intelligent Browsers 272
11.6.2 Common Platforms 272
11.6.3 Computational Theories of Meaning 273
11.6.4 Utilization of Intertextualities 274
11.6.5 Analyzing Networked Text 274
11.6.6 Natural Interfaces 275

Chapter 12 • Reliability 277

12.1 Why Reliability? 277
12.2 Reliability Designs 280
12.2.1 Types of Agreement; Types of Reliability 280
12.2.2 Conditions for Generating Reliability Data 283
12.2.3 Reliability Data 285
12.3 Agreement on Coding Predefined Units 289
12.3.1 The Direct Path to $_c\alpha$-Agreement 293
12.3.2 The Path to $_c\alpha$ via Coincidence Matrices 296
12.3.3 Difference Functions for Data With Various Metrics 299
12.3.4 Some Typical Examples and Important Properties 306
12.3.4.1 Binary Data, None Missing 306
12.3.4.2 Coincidences and Contingencies 308
12.3.4.3 $_c\alpha$'s Decomposition of Observed Coincidences 310
12.3.4.4 Two Observers, Several Nominal Values, No Missing Data 310
12.3.4.5 Uninformative Values in Ordered Variables 312
12.3.4.6 Trading Information for Reliability 314
12.3.5 Some Contested Coefficients and Correspondences 315
12.4 Accuracy, Surrogacy, and the Decisiveness of Majorities 323
12.4.1 Accuracy of or Surrogacy for Crowd Members' Judgments 325
12.4.2 Accuracy of or Surrogacy for Majorities or Aggregates 326
12.4.3 The Decisiveness of Majorities or Aggregates 328
12.4.4 Numerical Examples 329
12.5 The Reliability of Text Mining and Information Retrieval 331
12.5.1 For Replications of Searches 334
12.5.2 For Replications of Judgments on the Relevance of Retrieval Results 336
12.6 Agreement on Unitizing and Coding Finite Continua 337
12.6.1 Reliability Data From Unitizing 338
12.6.2 Observed and Expected Coincidences 340
12.6.3 Agreement on the Partitions of a Continuum Into Unequal Segments 342
12.6.4 Agreement on the Distinction Between Relevant and Irrelevant Segments 343
12.6.5 Agreement on the Coding of Single-Valued Segments Independent of Unitizing 344
12.6.6 Agreement on Individual Values Among Valued Intersections 346
12.7 Agreement on Multi-Valued Coding 347
12.7.1 Two Reasons for Multi-Valued Coding 348
12.7.2 Reliability Data for Multi-Valued Coding 349
12.7.3 Multi-Valued Difference Functions 349

12.7.4 Disagreements for Multi-Valued Data 350
12.7.5 $_{c}\alpha$ for Multi-Valued Coding 351
12.8 Statistical Properties of α 351
12.8.1 Insufficient Variation 351
12.8.2 Statistical Significance 353
12.8.3 Sampling Considerations 354
12.8.4 Standards for the Reliability of Data 356

Chapter 13 • Validity 361

13.1 Validity Defined 361
13.2 A Typology for Validating Evidence 365
13.2.1 Sampling Validity 368
13.2.2 Semantic Validity 370
13.2.3 Structural Validity 376
13.2.4 Functional Validity 377
13.2.5 Correlative Validity 378
13.2.6 Predictive Validity 381

Chapter 14 • A Practical Guide 383

14.1 Designing an Analysis 384
14.1.1 Text-Driven Analyses 384
14.1.2 Problem-Driven Analyses 386
14.1.2.1 Formulating Research Questions 387
14.1.2.2 Ascertaining Stable Correlations (With the Research Questions) 388
14.1.2.3 Locating Relevant Texts 390
14.1.2.4 Defining and Identifying Sampling Units Among Relevant Texts 392
14.1.2.5 Sampling a Sufficiently Large Number of These Units 393
14.1.2.6 Developing Appropriate Coding Categories and Recording Instructions 394
14.1.2.7 Selecting Appropriate Analytical Procedures 395
14.1.2.8 Adopting Standards for the Reliability of Generated Data and Statistical Significance Levels for the Results 396
14.1.2.9 Allocating Resources 397
14.1.3 Method-Driven Analyses 398
14.2 Writing a Research Proposal 400
14.2.1 Rhetorical Function 400
14.2.2 Contractual Function 401
14.2.3 Outline for a Research Proposal 402
14.3 Applying the Research Design 403
14.4 Narrating the Results 404
14.4.1 Outline for a Research Report 405

Glossary 407

References 414

Index 433

About the Author 453

PREFACE TO THE FOURTH EDITION

Content analysis is one of the most important research techniques in the social sciences. It acknowledges that society is enacted in talk, texts, and other modalities of communication and that understanding social phenomena cannot be achieved without understanding how language operates in the social world. Content analysts inquire into social phenomena by treating data not as physical events but as communications that are created and disseminated to be seen, read, interpreted, enacted, and reflected upon according to the meanings they have for their recipients. Interpreting communications as texts in the contexts of their social uses distinguishes content analysis from other empirical methods of inquiry.

Investigative methods in the natural sciences, by contrast, rarely are concerned with meanings, contents, intentions, references, communications, and what they do. Natural scientists may have good reasons to presume that the objects of their attention are not aware of how they are being observed and do not speak a language to articulate how they want to appear, but they also rarely reflect on how their own writings shape their conception of nature and what it allows them to do or prevents them from studying. The conception of nature as preexisting and independent of scientific attention is convenient for it relieves natural scientists from being held accountability for the realities that their theories construct, but it also prevents them from contributing to understanding social phenomena that arise in how they are described and enacted. Where social scientists adopt natural scientific methods of inquiry, it is the epistemology inscribed in these methods that prevents them from addressing what matters in the social life of people, what motivated and inspires them, and how they enact what they know. Certainly, content analysis is not the only research method that takes meanings seriously, but it is a method that has the additional advantage of being unobtrusive—not affecting the sources or receivers of what is being analyzed—able to process large volumes of data—providing insights that exceed the comprehension of ordinary readers of textual matter—and its findings can mediate between individuals and challenge the operation of institutions.

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 measurements provided the only evidence that counted (Lasswell, 1949/1965b), or they could redirect their attention to social phenomena that are both generated by and constituted in texts and images and, hence, needed to be understood through their written and pictorial manifestations. The latter motivated the first edition, and I am pleased to say that the epistemology and logic I developed therein have survived critical assessments and the challenges posed by the radical transformations that the information revolution and social media have introduced into contemporary social

practices. The widespread use of electronic, and hence computer-readable, texts concerning virtually everything that matters to us as social beings has moved content analysis, particularly computer-aided text analysis, into the center of how society examines itself.

In the 1970s, content analysis was a research method that had entered the psychological and social sciences but was used mainly by journalists and communication researchers. During that time, the amount of human effort required to collect, transcribe, and code textual data made content analysis a time-consuming, labor-intensive, and often costly effort that limited the scope of what it could address. However, it became an efficient alternative to public opinion research and a method of tracking markets, political influence, and emerging ideas. It was used to detect biases in reporting of facts, settle legal disputes, understand how institutions establish themselves in the texts they produce, and explore the mind of individuals through what they said or wrote.

The first edition of *Content Analysis* reported numerous analytical innovations but barely met the challenges of the new era of electronic communication. It taught a generation of social scientists about the methodology of this research technique. It soon became the leading text on the subject and was translated into Italian, Persian, Japanese, Spanish, Indonesian, and Hungarian.

Although the outline of the second, 2004 edition of *Content Analysis* remained essentially unchanged from that of the first, the book clarified numerous methodological issues and responded to the technique's latest challenges. All chapters were substantially rewritten, addressing developments that had taken place since 1980, especially concerning computer-aided text analysis. It added a practical guide that incorporated input from my students, colleagues, and own practical and consulting experiences in various academic and commercial research projects. In 2004, it received the International Communication Association's Fellows Book Award for its lasting contribution to the advancement of communication studies.

To broaden the range of examples and recommend solutions to practical problems, Mary Angela Bock and I coedited *The Content Analysis Reader* (2009). This reader was meant to be a companion to the methodologically oriented *Content Analysis*. I have been using the *Reader* with considerable success, heard from others about its added benefits, and recommend it to students, teachers, and practitioners of content analysis.

The third, 2013 edition of *Content Analysis* continued the careful and critical examination of how texts and images can support inferences about phenomena not accessible to independent observations and updated several chapters, notably on computer aids. It also added a convenient glossary. In 2014, it appeared in its Chinese translation. In 2017, my work on content analysis was honored by receiving the first methodological innovation award from the Mass Communication Division of the International Communication Association.

The outline of this fourth edition is kept essentially as it was before. It corrects a previously published error in the formula for computing the reliability of unitizations. Its discovery by a group of French scholars is just one example of the many fruitful collaborations that the book encouraged. It includes a section on crowdsourcing and expands Krippendorff's Alpha in several dimensions: for assessing the reliability of text retrievals, crowdcoding, and of multi-valued coding of predefined units as well as

of unitized textual continua. And it points to software among many versions known to be correct.

I thank my students at the University of Pennsylvania's Annenberg School for Communication for their interest and for giving me useful feedback and my colleagues for presenting me with the challenging methodological problems of their content analyses. I would also like to thank numerous readers of the earlier editions—both students and practicing content analysts—for sharing their comments and recommendations and Sage Publications for its continuing support of content analysis literature.

During the 38 years of publications of *Content Analysis*, its three previous editions have reached an enormous audience. It has been widely adopted as a text in social science, humanities, and business curricula, served researchers as a guide for the design and execution of large and small content analyses, and provided epistemologically sound standards 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. I hope this fourth edition will withstand critical examinations and continue to advance scholarship based on texts, images, and communications in all fields that acknowledge to operate within language.

—Klaus Krippendorff

Professor of Communication and
Gregory Bateson Professor Emeritus, for
Cybernetics, Language, and Culture

The Annenberg School for Communication,
University of Pennsylvania

INTRODUCTION

The term *content analysis* is about 70 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 six conceptual revolutions, as described below, and it probably is not stopping with these:

- The idea of *messages*: the early awareness not only that verbal discourse is portable when written but also 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 voice mail or say that a message was meaningful (full of meanings) or meaningless (void of meanings).
- The idea of *channels:* the awareness of 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 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 that it establishes relations between senders and receivers, enables the negotiation of complementary roles, constitutes social structures, and communities whose members have a sense of knowing each other. This awareness has ancient roots in the conception of dialogue but has entered current awareness through the

limitations experienced in the use of mass communication. By producing and disseminating identical messages—news and entertainment—to everyone, the mass media promised to be an agent of sharing, of building publics, ideally beyond borders. Modeled on the success of mass production and ever-expanding markets for useful goods, the mass media also made us aware of where this one-way model failed: making people aware of often amplifying social differences, limiting face-to-face conversations, curtailing public deliberations, and addressing only commercially viable issues. In U.S. culture, mass-media technology has become synonymous with progress, and the lack of communication is often blamed for most interpersonal problems or national conflicts.

- 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 largely defied theoretical accounts. 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 near universal participation.
- The idea of *computation:* the awareness of the increasing replacement of communicators by algorithms. Messages may be generated by robots and addressed to computers, whether it involves the use of automatic teller machines, online shopping, robo-calls, or computer-generated messages on the Internet. This changes the conception of messages as understandable by senders and receivers to what has computable consequences. Letter writing is being replaced by e-mailing, and interpersonal relations are shaped by the platforms that mediate them. Computation has expanded access to textual matter—consider the amazing power of search engines, the individual ability to travel to distant places virtually—but it precludes computationally unavailable experiences. Computation not only has made social life more variable but also has introduced complexities that easily escape understanding.
- The idea that reality is *discursively co-constructed*: the recognition that we live together in language, can never think alone (Sloman & Fernbach, 2017), and constitute our institutions as interdependent realities through what we say to and do with each other. Wittgenstein's (1958) notion of language as consisting of numerous language games that coordinate actions in the world comes to mind. The empirical fact that almost all innovations are the product of collaborations, that even scientific accomplishments arise and are vetted in discourse communities renders texts as actively supporting or challenging social realities.

This rather sketchy history of communication suggests that researchers who are concerned with texts can no longer treat them as symbolic or representational, 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 value (Krippendorff, 2017): *content,* the *what*

of a communication, something that authors are conceived of as *entering* into the texts they are producing, something that is *preserved* in texts while *shipped* to remote receivers, who are expected to *remove* its *content* as intended and henceforth *share it* with others through still other containers. This bizarre notion of message content renders authors as authorities of what messages contain and to the conception of *content analysts* as experts, able to objectively account for what messages *really contain* or are *intended to convey,* distinct from ordinary and usually inaccurate readings by less qualified readers.

The absence of authors, especially historical ones, makes the use of this container metaphor nonverifiable. Mass media messages tend to replace individual authors by institutional constructs, and their effects are more important than what someone may have intended. Trust in them has become trust in the technology of transmission. The mass media coordinate the lives of their audience members. Their practices challenge 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, textual data archives of social media—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 presumed target of the research effort—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 unflinchingly oppose the naive and misleading entailments of this still pervasive 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 microlevel 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. A novel solution of the human problem of analyzing increasing volumes of textual data is to *outsource the coding of textual matter to crowds who can interpret data reliably.* The methods to cope with this practice are currently at their infancy. Developing them moves content analysis closer to other social science methods such as survey research. Economy and speed are noticeable benefits of outsourcing the coding of textual data, but reliability and depth are still issues to be improved upon.

- *The large volumes of electronically available data call for computer aids.* Such aids need to convert large bodies of electronic text into more manageable representations that preserve the ability to answer the research questions that content analysts seek to derive from available data. Currently, two competing approaches are pursued to handle overwhelming volumes of data. One is to develop text analysis software capable of computing complex statistics of the character strings of text or pixelized images. However, exactly how such software transforms raw texts into abstract representations is often uncertain and the subject of exaggerated claims, for example, being content extractors or promising to mine meanings. The other approach is to develop less ambitious computer aids that merely replace the labor-intensive and recurrent parts of more complex content analyses. For example, qualitative data analysis software tends to allow recursive coding decisions to be automatically extended to large documents. Software that could serve as aids to parts of content analytical steps requires some transparency, enabling analysts to assess their appropriateness by tracing and evaluating the transformations that texts are undergoing in the process.

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 procedures; to evaluate the performance of individual techniques; and to estimate the likelihood of particular research designs to contribute to knowledge. All researchers must become proficient in defining the terms of their analysis and justify the analytical steps taken to a skeptical friend or critical colleague. Methodology provides a language for talking about the process of research, not about subject matters. 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, for example, of the *Iliad* before Homer put it in writing. It was only a century ago that the historian Leopold von Ranke 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 critics claim that each content analysis is unique, possibly by focusing mainly on its subject matter, I would argue that all content analyses share a procedural logic that can be examined and is used to justify their procedures in terms of socially acceptable criteria. What content analyses have in common 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, as citizens, we can hardly 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 or build on them.

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," "sampling," "recording/coding" in terms of formal "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 undertaken 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 it suits their research interests. If it does, they might want to 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 undertaking a larger project. 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 content analyzing text is not a mechanical task and neither is designing of a project. 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 and recording content analysis data. Attention to their problems is particularly important when data or texts are unwieldy, complex, and deserving of careful attention. 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. I like to mention *The Content Analysis Reader* (Krippendorff & Bock, 2009), which provides many more examples of solutions to practical problems than this book could discuss in sufficient length. 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 the conceptual framework they seek to operationalize enriches an analysis. Such readers might well look into Chapter 11, on computer aids, to gain an alternative perspective on manual unitizing and coding.

Readers who have already undertaken content analyses or similar text-based research will discover in this book alternative paths for inquiries of texts and a vocabulary that can be used in deliberating about what is involved in analyzing texts—unlike observational studies of naturalistic phenomena—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 (on the conceptual foundations), 12, and 13 (on reliability and validity) especially applicable to evaluative endeavors.

While this book may serve as a handbook for various practitioners, it grew out of my experiences in teaching courses, seminars, and workshops in content analysis. In the context of education, 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 11 more important than Chapter 7, on recording/coding, and may omit Chapter 12, on reliability, which is not a problem for software applications; but they ought to consider the easily ignored validity of computer uses, discussed in Chapter 13. 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 what they offer.

Finally, for me, the book will have achieved its purpose if it helps to make the newly available wealth of electronic texts accessible to systematic analysis, if it improves the social significance of research in the humanities and the social sciences, and if it encourages developments beyond the methods discussed therein. Understanding the realities that social actors construct with the textual matter they produce and transmit is an ongoing challenge.