

Content Analysis

An Introduction to Its Methodology

Fourth Edition

Klaus Krippendorff
Annenberg School for Communication
University of Pennsylvania



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

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 <i>Decision Schemes</i>	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 <i>Chains</i>	168
8.5.2 <i>Recursions</i>	169
8.5.3 <i>Cubes</i>	169
8.5.4 <i>Trees</i>	171
8.5.5 <i>Multi-Valued Sets</i>	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 ζ_α -Agreement	293
12.3.2 The Path to ζ_α 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 ζ_α '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_{α} 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 <i>Formulating Research Questions</i>	387
14.1.2.2 <i>Ascertaining Stable Correlations (With the Research Questions)</i>	388
14.1.2.3 <i>Locating Relevant Texts</i>	390
14.1.2.4 <i>Defining and Identifying Sampling Units Among Relevant Texts</i>	392
14.1.2.5 <i>Sampling a Sufficiently Large Number of These Units</i>	393
14.1.2.6 <i>Developing Appropriate Coding Categories and Recording Instructions</i>	394
14.1.2.7 <i>Selecting Appropriate Analytical Procedures</i>	395
14.1.2.8 <i>Adopting Standards for the Reliability of Generated Data and Statistical Significance Levels for the Results</i>	396
14.1.2.9 <i>Allocating Resources</i>	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