From Words to Numbers

This book illustrates a set of tools – story grammars, relational data models, and network models – that can be profitably used for the collection, organization, and analysis of narrative data in sociohistorical research. A story grammar, or subject-action-object and their modifiers, is the linguistic tool the author uses to structure narrative for the purpose of collecting data on protest events. Relational database models make such complex data collection schemes practically feasible in a computer environment. Finally, network models are a statistical tool best suited to analyze this type of data. Driven by the metaphors of the journal (from . . . to) and the alchemy (words into numbers), the book leads the reader through a number of paths, from substantive to methodological issues, across time and disciplines: sociology, linguistics, literary criticism, history, statistics, computer science, philosophy, cognitive psychology, and political science. The book mitigates its quest for rigor in the social sciences with a subtle irony for that quest.

Roberto Franzosi obtained an Honors BA in Literature at the University of Genoa (Italy) (1975) and a Ph.D. in Sociology at Johns Hopkins University (1981). After spending a year as a postdoctoral Fellow at the University of Michigan, he taught for several years in the Sociology Department at the University of Wisconsin-Madison. Between 1995 and 1999, Franzosi held a lectureship in Sociology at the University of Oxford, with an official fellowship at Trinity College. He took the post of head of the department in Sociology at the University of Reading in 1999. He has served as consulting editor for the American Journal of Sociology and is currently a managing editor for the Journal of Historical Sociology. The substantive and methodological study of social protest has been at the heart of Franzosi’s research agenda with several published articles and the book The Puzzles of Strikes (Cambridge University Press, 1995).
Structural Analysis in the Social Sciences 22
Mark Granovetter, editor

The *Structural Analysis in the Social Sciences* series presents approaches that explain social behavior and institutions by reference to relationships between such concrete entities as persons and organizations. This contrasts with at least four other popular strategies: (1) reductionist attempts at explanation by focusing on individuals alone; (2) explanations stressing the causal primacy of ideas, values, and cognitions; (3) technological and material determinism; (4) explanations using “variables” as the main analytical concept, as in “structural equation” models, where the structure connects variables rather than actual social entities.

An important example of structural analysis is the “social network” approach. However, the series also features social science theory and research that is not framed explicitly in network terms, but stresses the importance of relationships rather than the atomization of reductionism or the determinism of ideas, technology, or material conditions. Such efforts typically deal with the complex balance between structure and agency, increasingly a key issue in the human sciences. Examples of the structural approach are scattered across many disciplines, and it is the goal of the *Structural Analysis* series to expose this very fruitful style of analysis to a wider public by bringing all the approaches together under a single rubric.

*Other books in the series:*

1. Mark S. Mizruchi and Michael Schwartz, eds., *Intercorporate Relations: The Structural Analysis of Business*
2. Barry Wellman and S. D. Berkowitz, eds., *Social Structures: A Network Approach*
3. Ronald L. Breiger, ed., *Social Mobility and Social Structure*
4. David Knoke, *Political Networks: The Structural Perspective*
5. John L. Campbell, J. Rogers Hollingsworth and Leon N. Lindberg, eds., *Governance of the American Economy*
6. Kyriakos Kontopoulou, *The Logics of Social Structure*
7. Philippa Pattison, *Algebraic Models for Social Networks*
8. Stanley Wasserman and Katherine Faust, *Social Network Analysis: Methods and Applications*
9. Gary Herrigel, *Industrial Constructions: The Sources of German Industrial Power*
10. Philepe Bourgois, *In Search of Respect: Selling Crack in El Barrio*
11. Per Hage and Frank Harary, *Island Networks: Communication, Kinship, and Classification Structures in Oceania*
12. Thomas Schweizer and Douglas R. White, eds., *Kinship, Networks and Exchange*

*Continued on page following the Index*
From Words to Numbers

Narrative, Data, and Social Science

ROBERTO FRANZOSI

University of Reading
Andreae patri carissimo, in memoriam.
Sit tibi terra levis.
Non chiederci la parola che squadri da ogni lato
l’animo nostro informe, e a lettere di fuoco
lo dichiari e risplenda come un croco
perduto in mezzo a un polveroso prato.

Ah l’uomo che se ne va sicuro,
agli altri ed a se stesso amico,
e l’ombra sua non cura che la canicola
stampa sopra uno scalcinato muro!

Non domandarci la formula che mondi possa aprirti,
si qualche storia sillaba e secca come un ramo.
Codesto solo oggi possiamo dirti,
ciò che non siamo, ciò che non vogliamo.

Eugenio Montale (from Ossi di seppia)

Do not ask us for the word that will square off our shapeless soul from every angle, and proclaim it in characters of fire and shine like a crocus lost in a dusty field. Ah, the man who goes through life with confidence, a friend to others and to himself, unconcerned with his shadow cast by dog days on a flaking wall! Do not ask us for the formula that will open worlds for you, but only some twisted syllables as dry as a branch. Today, we can tell you only this, what we are not, what we do not want (Montale, 1965).
Contents

List of tables  page  xiii
List of figures  xv
Acknowledgments  xvii

Before
The Ant and the Cicada  1
“It Takes Two to Tango”: In Search of the Actor  3
The Lemon and the Grapefruit  4
Of Journeys and Alchemies  5
“The Law of Genre”  8
“Methodological Pestilence”  11
An Antiquarian in a Modern World  14
Breaking the Law  18
Of Prefaces and Introductions: Guidebook for the Journey  20

Part I  From Words to Numbers: A Journey

Prologue  27

Chapter 1  “In the Beginning Was the Word”  29
“In the Beginning Was the Word”  29
Content Analysis: “Objective, Systematic, and Quantitative”  32
From Event Counts to Event Characteristics  35
Turning to Linguistics for Help  40
Syntax Grammar  41
Semantic Text Grammar  43
Psychologists’ Story Grammars  45
Contents

A Story Grammar for Sociohistorical Research 47
Examples of Coding 52
Running into Problems with Story Grammars 54
Narrative Discourse 55
What Have We Gained? 59

Chapter 2  Ars Memoriae 62
The Artificial Art of Memory: The Old and the New 62
Database Design for a Story Grammar 66
Computer-Assisted Content Analysis 69
  Basic Features of the Software 70
  Some Lessons from Cognitive Psychology 72
  Reliability versus Efficiency 73
On Sampling: More on Reliability versus Efficiency 73
And More Lessons from Linguistics: Text Schemata 74
Data Verification 76
Navigating a Relational Database 79
Toy Examples and “Real” Data: The Test of Large Data Sets 81
  From the “Red Years” to the “Black Years” (Italy, 1919–1922) 82
  Simple Counts Tell Complex Stories 82
Stitching the Thread 89
Further Advantages of Story Grammars 91
Scientists’ Bitter Fate 92

Chapter 3  “Everything Is Number” 98
“Friends of Friends”: Grammars and Network Models 98
Mobilization and Counter-mobilization Processes (1919–1922):
  Shifting Networks of Interaction 100
“Everything Is Number” 109
Of Thin Threads and Other Threads 113
And Yet, What Do We Know? 118
In the Linguists’ Tool Bag: Theta Theory 122
Of Beauty (and Science) 124

Journeys 131
“A March it Was” 131
Pilgrims on the Road 132
Voyages of Discovery 138
Ways of Seeing 142
Silence and Emphasis 147
“Journeys . . . Magic Caskets Full of Dreamlike Promises” 154
Contents

Part II  Looking Back: What’s in the Numbers?

Prologue 157

Chapter 4  The Word and the World 161

Founding Fathers and Disciplinary Mottos 161
The “Science of Sociology” 161
“Wie es eigentlich gewesen” 164
The Newspaper “Black Box” 167
Why, Then, Newspapers? Social Scientists’ Line of Defense 171
Opening the Box: Behind the Surface of Our 1919–1922 Data 173
Silence and Emphasis Revisited 175
How to Go about Validating Data: A Decalogue of Rules 177
Further Rules: Selecting the Newspaper(s) 180
Barking up the Wrong Tree: Reliability versus Validity 180
Data: “Given” or Constructed? 183
Content and Form 187

Chapter 5  “A World of Words” 193

Back to the Linguist’s World 193
Stories 194
An Oxford Story 194
A Children’s Story 195
The Moral of the Stories: “Where Is Anna Giulia?” 196
“The Body and its Shadow”: Play, Aesthetics, and Language Games 199
The Language of Science 202
Poetry and Science 207
Heart and Mind 211
Dialoghi Italiani 214
Confessions of an Ass 216
Content Analysis Myths (Back to Silence and Emphasis) 219
Debunking the Myths: What Is Meaning? 222
True Believers 226
Modern Alchemists: What We Do with the Numbers 229
What’s to Be Done? Alchemists’ Recommendations to the Novice 232

Chapter 6  Journeys’ Ends 237

Sta Viator 237
From Words to Numbers 238
From Thin Explanations to “Thick Descriptions”? 238
From Variables to Actors 240
Contents

xii  From Synchronic to Diachronic Time  242
From Statistical Causality to Narrative Sequences  243
From Structures to Events  245
Unexpected Crossings  247
Sociology Meets History  248
The Centrality of the Verb (Linguistics and Sociology)  252
Sociology as the Study of Social Relations  255
Georg Simmel (1858–1918): “The Stranger in the Academy”  256
Von Wiese’s “Specificum Sociologicum”  260
Simmel’s Cash Among the Heirs  263
“Fatti Maschi, Parole Femmine”  264
The Feminist Critique  266
Memorabilia  270

Tales of Measurement  274
Tales of Horror  274
Tales of Pressure  276
Social Science Tales  279
Wars of Words  282
What Have We Lost?  285
“Garçon, l’addition!” Counting What?  287
Classification and the Numbers  288
Classification in Practice: The Secret of Secrets  293
The Words (and the Worlds) of Science and Metaphor  295
Tales of Time and Space  298
Faust and His World: What Was in the Beginning?  301

And After  309
“So Long as We Get Somewhere”  309
Poupinel and His World: In the Beginning Was Meaning  312
On the Road, Again  315
“Know Thyself”: Notes on Reflexive Sociology  318
Origin and Originality: Limits to Innovation  324
“Murmure Against Me”  328
Farewell to the Reader  332

Appendix  339
Notes  343
References  399
Index  445
List of Tables

1.1. Rough number of published studies based on content analysis (five-year intervals, 1921–1950)  page 34
2.1. A relational database representation of a story grammar  68
2.2. The skeleton triplets of the Romeo lockout article in relational format  81
2.3. Frequency distribution of articles, disputes, and triplets in the 1919–1922 database  84
2.4. Frequency distribution of actors (subjects and objects) in 1919 (number of disputes)  84
2.5. Frequency distribution of most common actors (number of disputes)  84
2.6. Frequency distribution of actions in 1919 (number of disputes)  86
2.7. Frequency distribution of most common actions (number of disputes)  86
2.8. Frequency distribution of most common demands (number of disputes)  86
2.9. Frequency distribution of the most common actions performed by workers (number of disputes)  88
2.10. Frequency distribution of the actor “Fascists” (number of disputes)  88
2.11. Frequency distribution of the actions performed by Fascists (number of disputes)  88
2.12. Frequency distribution of the actors with whom Fascists “clash” (number of disputes)  89
3.1. Factor analysis of eighteen types of collective actions in the 1919–1922 database  114
List of Tables

3.2. Factor analysis of eighteen types of collective actions in the 1919–1922 database (rotated factor matrix) 115
3.3. Least-squares estimates of a regression model of fascist violence 117
4.1. Distribution of disputes from Il Lavoro by location of occurrence 174
List of Figures

2.1. Van Dijk’s schema of news discourse ........................................ page 75
2.2. Plot of the yearly number of strikers (1879–1922) ....................... 83
2.3. Plot of the monthly number of strikers (1919–1922) ................. 83
3.1. Network graphs for the relations of conflict and facilitation for the Romeo Factory Lockout article from Il Lavoro ................................. 101
3.2. Network graphs for the sphere of action of communication .......... 102
3.3. Network graphs for the sphere of action of request .................... 103
3.4. Network graphs for the sphere of action of conflict .................... 104
3.5. Network graphs for the sphere of action of violence .................. 105
3.6. Network graphs for the sphere of action of control .................... 107
3.7. Network graphs for the sphere of action of facilitation ................ 108
4.1. Plots of the monthly number of triplets with workers or Fascists as subjects or objects .......................................................... 175
Acknowledgments

Long projects inevitably end up with long lists of accumulated debts to people and institutions. This project is no exception. I developed the ideas at the heart of this book at the Center for Research on Social Organizations at the University of Michigan, where I spent the 1981–82 academic year as a postdoctoral Fellow. Charles Tilly, director of CRSO, and Nancy Horn, a graduate student working on Tilly’s Great Britain Study, were very generous with their time in clarifying the inner workings of Tilly’s approach to coding text for historical research. Charles Tilly’s influence on my work did not stop with the year spent at CRSO; neither did it confine itself to methodological issues. What has lived in me through the years is the legacy of Tilly’s passion for historical questions that can be addressed through rigorous methods, an unfailing excitement for the new and the innovative.

A grant from Ente Nazionale Idrocarburi (ENI) allowed me to spend a year in Rome after Ann Arbor exploring new ways of collecting text data, taking Tilly’s scheme as a point of departure. Gustavo DeSantis played a crucial role in making it happen. Anna Maria Ventrella, director of ISVET, one of ENI’s research institutes, gave warm hospitality to my research team, as did all the members of the institute. Annella Centis, one of the team members, stuck with the project till the very end. Her help was invaluable.

In 1986, I started the long and bumpy ride of data collection in Genoa, Italy. If the beginnings were exciting, by the time I finished data collection in 1990, I had run out of steam, amidst financial and operational difficulties. Throughout those years, Giorgio Sola did more than one can reasonably expect of a friend and a colleague. Many students from the University of Genoa worked on the project as coders; they are too numerous for me to thank individually. Alberto Devoto, of the Istiituto Nazionale di Fisica Nucleare of the University of Cagliari, and Corrado Salvo, of the Department of Physics of the University of Genoa, provided the VAX accounts that made possible computer communications between Madison and Genoa. Between 1995 and 1997, Piero Squillante, Mauro Giorgetti, and Daniela Mariani worked on data cleaning and
data analysis. I am particularly grateful to Pier Paolo Mudu for his role in data analysis.

Several institutions provided research funds: the Graduate School of the University of Wisconsin, Formez, Confindustria, Intersind, Consiglio Nazionale delle Ricerche (Progetto strategico “Il conflitto e le relazioni di lavoro nel prossimo decennio”), and the National Science Foundation (SES-8511632 and SBR 94-11739). The following people played key roles in securing the grants: William Bainbridge, Mario Bruti, Gian Primo Cella, Piero Celli, Innocenzo Cipolletta, Giulio De Capraris, Benedetto de Cesaris, Enzo Grilli, Charles Manski, Enrico Micheli, Mario Micossi, Filippo Peschiera, Giuseppe Rosa, Carlo Sampietro, and Goffredo Zappa. In am particularly indebted to David Ward at the University of Wisconsin-Madison who gave me the initial funds (and confidence) to embark upon this long journey.

Thomas Mitchell, Shawn Pouchot, and José Carlos De Almeida Manganaro proved to be the most competent and most reliable programmers in the development of the data-entry software PC-ACE (Program for Computer-Assisted Coding of Events). Charles Palit and John Nitti at the University of Wisconsin, both involved in different projects based on “words” and on PC technology, provided constant and invaluable technical help and support (and so did my colleague Halliman Winsborough).

In 1986, I submitted my first manuscript on the methodological approach I had developed to Sociological Methodology. The comments I received from one of the reviewers indicated that I should fundamentally change my thoughts, away from artificial intelligence and closer to linguistics. Clifford Clogg, editor of the journal, strongly encouraged me to address the reviewer’s recommendations (“a foremost linguist in the country,” in Clogg’s words). Charles Scott at the University of Wisconsin-Madison guided my first steps through an unfamiliar terrain: linguistics. To that anonymous “foremost linguist” goes my deepest gratitude. Perhaps, s/he may never know how deeply her/his review changed the course of my intellectual and professional trajectory. Thanks to you as well, Clifford Clogg. Tibi quoque sit terra levis. Peter Marsden, successor to Clifford Clogg at the helm of Sociological Methodology, and J. Scott Long, editor of Sociological Methods and Research, were just as supportive with further publications on my method.

Several years were still to pass between the time I submitted a collection of published papers to Cambridge University Press (fall 1992) and the time I submitted the completed manuscript for final review (June 2000). Many things came in between – the writing of The Puzzle of Strikes, three successive moves of institutions, houses, and even countries, the birth of our daughter, Anna Giulia, in December 1995, and, more to the point, a change in my intellectual trajectory that forced upon me a slow and painful rethinking of my work along interdisciplinary lines.

Many people played key roles in providing guidance and insights in this wild-goose chase across disciplinary boundaries. I list them here in alphabetical order: Ronald Aminzade, Alberto Devoto, James Jaffee, Donald Levine,
Acknowledgments

Diana Lewis, John Markoff, Larry Scanlon, Biljana Scott, Charles Tilly, Louise Tilly, and Harrison White. Harry Eckstein, Edward McDill, James Rule, Rudy Rummel, and Raymond Tanter were kind enough to illuminate aspects of the history of the process of quantification in which they were involved. But many more than I can list offered occasional references and suggestions. Several research assistants at Oxford helped me with the extensive literature: Michael Elliot, Michelle Jackson, Richard Lim, Shepley Orr, and Daniel Viehoff. Desmond King, Nader Sohrabi, Charles Tilly, and Gavin Williams provided helpful comments on chapters of the manuscript. “Sraffa had the power to force Wittgenstein to revise, not this or that point, but his whole perspective,” Ray Monk tells us (1991, p. 260). I had a few Sraffa of my own who, sometimes with off-hand comments, pushed me to explore new directions or substantially revise the manuscript: Richard Gipps, Mark Greengrass (to whom I owe the metaphor of the journey), Daniel Schwartz, Giorgio Sola, and Daniel Spizzo.

At the University of Oxford I found an exciting intellectual atmosphere that will be hard, if not altogether impossible, to replace. Michael Inwood, Desmond King, and Gavin Williams played deeply key roles in my intellectual development during those years. The Department of Applied Social Studies and Social Research provided a comfortable, friendly, and supportive atmosphere for work. At the Taylorian Library and Rhodes Library I spent many days and some wonderful summers researching and reading. There, Ms. Elisabeth Tabouis and Mr. Allan Lodge provided unfailing help and always with a smile. At Trinity College, I constantly broke the fundamental rule of convivial life: “Never talk about work.” Instead, at the senior common room lunches and dinners I picked everyone’s brain at every opportunity; the historians, the philosophers, the mathematicians, the classicists, and even the scientists, all had to put up with my questions. I lived a childish dream that close physical proximity with so many different scholars may somehow transfer their knowledge onto me. Alas, the only knowledge that I did acquire was a humble and unequivocal sense of the limits of my own knowledge. Many of those conversations simply sparked in me a lusty gusto for a good story. In many ways, this book is the result of the Trinity Senior Common Room. There, I most acutely experienced the tension of the “two cultures,” so deeply ingrained in my own biography. Besides Mike Inwood, I am deeply indebted to several of my Trinity fellows for their constant suggestions: Dinah Birch, Peter Brown, Clive Griffin, Martin Maiden, Jonathan Mallinson, Bede Rundle, and Brian Ward-Perkins.

Needless to say, the final product had little to do with the 1992 book proposal. The beginning of that final product goes back to January 1996 when I started writing in the wee hours of the night, in between nightly feeds of Anna Giulia – a labor of serenity and love – born 21-12 (1995) at 21:12 in Oxford (the magic of numbers!). It continued in earnest till 1999, when I took a job as head of department at the University of Reading. The demands of rebuilding a department fallen under receivership took a heavy personal and professional toll. Lazarsfeld assures us that “educational innovations are, by definition, intellectual as well as administrative tasks” (Lazarsfeld, 1993, p. 269). Then, it was
Acknowledgments

not in vain. For sure, a free hand in setting up new undergraduate and graduate curricula and to hire several new colleagues forced me to bring into sharp focus what I fundamentally believe sociology to be and how to best convey that view to students and colleagues. Nonetheless, the manuscript laid untouched in my laptop for a few more years. I am grateful to Mark Granovetter, editor of the series “Structural Analysis in the Social Sciences,” who saw the beginning of my journey, for believing in this travel diary, and to Lewis Bateman, the social science editor at Cambridge University Press, for pushing me to finally let the manuscript go. I am also grateful to Richard Laver for his excellent work with the index, Paul Luna for typesetting advice, and Sophie Doyle for checking and rechecking everything in the final stages of production. The camera-ready copy of the book was printed at the University of Trento where, in the fall of 2003, I resumed data collection on the 1919–22 period, with research grants from the University of Reading and the University of Trento.

This journey from words to numbers has been long and difficult. It has been both personally and professionally taxing, with some painfully humiliating moments. Several friends never gave up on me through it all: Alberto Devoto, Mauro La Noce, Alberto Palloni, Giorgio Sola, Charles Tilly, and Erik Wright. Their friendship and belief in my journey was crucial for me to finally reach a safe harbor. From 1997–1998, I increasingly retreated within myself and within the world of this book. Isolation bought me freedom but delivered grief. The family paid a high price for my estrangement. My wife, Svetlana Kropp, has been a constant source of both professional and personal support throughout most of these years, and certainly, throughout the hardest ones. She worked as a supervisor on the project in Genoa in 1990. She read more than her share of my work. She helped with data cleaning. She carried out a great deal of the early statistical analyses. She also did what no other had to do: She put up with me through it all. Mea culpa.

The following individuals and publishers have kindly given me permission to reprint material from the publications listed below for the poems and epigraphs used in this book:


Acknowledgments


xxii  Acknowledgments


