Content Analysis: Objective, Systematic, and Quantitative Description of Content

Roberto Franzosi

By the time of the publication of the first general textbook in content analysis in 1952 (Berelson's *Content Analysis in Communication Research*), the basic ingredients of the new methodology had all been worked out. And so had the rhetorical format of introducing the technique through a list of definitions by various authors – we will find that same format in later textbooks from Holsti to Krippendorff, and in various accounts of the development of content analysis from Shapiro and Markoff to Franzosi.[1] Sticking to this format, the following early definitions[2] leave no doubt about the quantitative nature of the technique:

> "[the method of] *quantitative* content analysis … consists of tabulating the occurrences of content units …"
>
> "Content analysis … attempts to characterize the meanings in a given body of discourse in a *systematic* and *quantitative* fashion. Content analysis is the *statistical* semantics of political discourse. … Content analysis aims at *statistical* formulations, directed toward empirical problems … its statistical character [is] one of its most distinctive attributes."
>
> "A distinguishing characteristic of content analysis … is its *quantitative* aspect."

Indeed, content analysis was born as a quantitative technique. Harold D. Lasswell, Lerner, and de Sola Pool (1952: 45; emphasis added), the father founders of the technique, put it in these words: "There is clearly no reason for content analysis unless the question one wants answered is *quantitative*."[3]

Quantification was the result of frequencies obtained through counting.[4] In his statement of the state of the art of content analysis, by 1959 standards, de Sola Pool (1959: 195) wrote: "Counting frequencies was the main activity of content analysts in the 1930s and 1940s. Indeed, for many people that is how content analysis was defined. Berelson's book minus one chapter is almost wholly devoted to such frequency counts. Harold D. Lasswell's content analyses were frequency counts of symbols …."[5] But

"why be quantitative?" That question provides the title to an introductory chapter in Lasswell et al.'s *Language of Politics. Studies in Quantitative Semantics* (and with which this collection opens). In addressing the question, Harold D. Lasswell makes clear his position in a choice between handling historical and social problems "with more exact methods" and "with precision" versus the alternative of a purely "qualitative, *impressionistic* and conjectural" approach (Lasswell, 1949: 47; emphasis added). By the time he opens his concluding paragraph with the summary question "Why, then, be quantitative about communication?" the reader has little doubt about Harold D. Lasswell's answer (1949: 52): "Because of the scientific and policy gains that can come of it."

Yet, while quantification is certainly a defining feature of content analysis, it was hardly the only (or perhaps even the main) one. A *systematic* and *objective* approach to measurement issues was no doubt just as important. In one of the earliest documents on content analysis, Waples and Berelson (1941: 2) wrote: "*Systematic* content analysis attempts to define more casual descriptions of the content, so as to show *objectively* the nature and relative strength of the stimuli applied to the reader or listener." Harold D. Lasswell himself insisted that (1942: 15; emphasis added):

> *Impressionistic* or systematic procedures may be used in describing communication content. When *impressionism* prevails, individuals characterize content without specifying the criteria that they use in making up their minds. They take no special precautions to check their judgments against those of other trained people in order to reduce the effect of whim and caprice. When *systematic* procedures are used the criteria of judgment are made as explicit as possible.

This anathema against "impressionism" would be picked up by others, close to Harold D. Lasswell. Kaplan compares content analysis to an "'*impressionistic*' method" of inquiry. Janis and Fadner write: "*Impressionistic* judgments suffice for broad classification of symbol data and description of gross temporal changes in the content of mass communications. … But if we wish to develop *precise hypotheses* concerning mass communications, there is a *need for quantitative analysis of symbols.*" Leites and de Sola Pool "contrast the 'objectivity' of content analysis with other more 'subjective' or '*impressionistic*' ways of talking about symbols."[6] Not surprisingly, Berelson, in his textbook, would define content analysis as: "a research technique for the *objective, systematic,* and *quantitative* description of the manifest content of communication."

Quantification, in this view, goes hand in hand with systematization, rigor, precision, and exactitude in definitions and measurements, with objectivity and replication of procedures and findings, in other words,

with a scientific approach to social science. And the goal of that scientific approach – to which content analysis is seen as an important methodological contributor – is to rigorously test hypotheses drawn from broader theoretical frameworks. "Content analysis marks the most promising direction along which the sciences of signs might proceed to *formulate and substantiate precise empirical hypotheses* … ."[7] Hypothesis testing features prominently in Leites and de Sola Pool's programmatic document "On Content Analysis" (1942: 21; emphasis added):

Four major functions of content analysis may be distinguished:

a. to produce increases in the degree of *confirmation of hypotheses* already generally presumed to be valid and definitive *disconfirmation of hypotheses* already generally presumed to be invalid.

b. To correct "optical illusions" which may be shared by most specialists.

c. To settle disagreements among specialists as the truth value of certain *propositions.*

d. To permit (1) the formulation and (2) the *testing of hypotheses* …

Indeed, the great "potentiality of content analysis" lies in hypothesis testing carried out in a systematic and scientific way (Janis, 1943: 429–430). And that potentiality is in addition to "the general advantage that is afforded by quantitative results, namely, the applicability of the probability calculus by use of inductive statistical tests of significance. If for no other reason than this, *content analysis is an invaluable scientific tool for the study of symbolic behavior*" (Janis, 1943: 430; emphasis added).

The overall aim of such a research program, as expressed by Harold D. Lasswell and his associates, was anything but trivially concerned with picky methodological questions. No. The aim was much farther reaching, as Harold D. Lasswell himself put it in one of the early documents (1941: 1, 12; emphasis added): "The technique of symbol analysis, properly applied, can provide us with insight into the lives of others by showing us what has come to their attention. … By means of symbol analysis [content analysis] it is clear we are able to *arrive at rather unambiguous descriptions of fundamental features of society.*"

Where did the Label "Content Analysis" Come from?

Writing in September 1942 (p. 1; emphasis added), Leites and de Sola Pool state: "During the *last few years* the term 'content analysis' has been increasingly used by students of symbolic aspects of society, particularly in connection with the studies and suggestions of Harold D. Lasswell." In November of that same year, Janis and Fadner (1942: 2; emphasis added) similarly write: "In *recent years* a number of studies … have employed

quantitative content analysis." A year later, Kaplan (1943: 230; first emphasis added) writes: "In *recent years* there has been (and is being) developed, especially by Harold D. Lasswell and his associates, a technique known as *content analysis.*"

It appears from these citations that by the end of 1942, the label "content analysis" had already been in use for at least a few years. It also appears that Harold D. Lasswell was a key figure in this development.[8] Certainly, the term "content" appears in many titles, even a few decades earlier than 1942 (in such expressions as content of newspapers, content of radio programs, editorial content, mass media content, content of mass communications, or communication content).[9] But the label "content analysis" does not appear in writing until 1940.[10]

In that year, Waples,[11] Berelson, and Bradshaw use the label in their book *What Reading Does to People* (1940).[12] A year later, in April 1941, Waples and Berelson use the label in the title of a typed document: "What the Voters Were Told (An Essay in *Content Analysis*)." Harold D. Lasswell himself, however, despite the wide acknowledgments for his role in the development of the label, shows some ambiguity in his choice of label.[13] In April 1941, he does use the label "content analysis" in the title of one of the internal documents of the *Experimental Division for the Study of War Time Communications,*[14] "The Technique of Symbol Analysis (*Content Analysis*)," but in parenthesis and alongside another main label, "symbol analysis." Still in 1949, in one of the first edited books that collected the early methodological and substantive contributions of the new technique of content analysis, Harold D. Lasswell and Leites prefer the title "Quantitative Semantics," rather than content analysis – a label still used by Dovring in her 1954 article "Quantitative Semantics in 18th Century Sweden." The publication, in 1952, of Berelson's *Content Analysis in Communication Research,* hailed as "the standard codification of the field" (de Sola Pool, 1959: 1), and of de Sola Pool's *Trends in Content Analysis,*[15] in 1959, clinched the label once and for all: Later textbooks would all have the standard label "content analysis" in their titles.[16]

Setting the Methodological Standards

Methodology was one of the main concerns of early contributors to the development of content analysis. Under the heading of "Take-Off Period," I have collected in Volume I some of these contributions, particularly those coming out of the *Experimental Division for the Study of War Time Communications* directed by Harold D. Lasswell. While some of these contributions were later published in revised form in journals or, especially, in the first book on content analysis (Lasswell et al., 1949), I have tried, to the extent possible, to use the original documents, published here for the first time.[17]

And among these methodological preoccupations, the design of appropriate coding schemes figured prominently.[18] For Harold D. Lasswell and his associates, coding scheme design should be strictly based on a theory of politics, on the "scientific understanding of politics,"[19] rather than on ad-hoc considerations (as typically found in the literature … still today!).[20] As Harold D. Lasswell wrote (*1942: 3): "Since by means of content analysis we provide data for a science of politics, our observations must be related to a systematic body of postulates, definitions, rules, hypotheses and laws." Content analysis plays an ancillary (empirical) role in this larger science of politics, and coding scheme design follows from this view of politics as science. "The stream of communication is made up of statements, and the key questions to ask about any statement are: What is said? Who says it? Who is affected, how? Symbol analysis [content analysis] is concerned with 'What is said'" (*Lasswell, 1941: 1).

Definitions beget definitions. What is a statement? Statements are basically collections of symbols and can be of three types: demand, identification, and fact (*Kaplan, 1943: 241).[21] These statements can be captured by two basic categories: direction and standards, where standard refers to a topic/theme addressed by a statement and direction refers to the positive, negative, or neutral reference to the standard.[22] Although the potential number of standards "is infinitely great" (Lasswell and Asoociates, 1942: 16), and the list provided in Harold D. Lasswell (*1941: 5–7; *1942: 27–29) is long (some twenty of them), "the categories of standard most frequently used are 'strength-weakness' and 'morality-immorality'."[23] Symbols as well, although potentially unlimited, in the narrow study of politics, can be conveniently grouped into a small set of "key symbols": "1. Persons (Roosevelt, Stalin, …), 2. Groups (Americans, Russians, …), 3. Agencies (Congress, Soviet, …), 4. Policies (War, Peace, …), 5. Participations (Enlist, Subscribe, …), and Ideas (Democracy, Nazism, …)."[24]

Don't be too optimistic about being able to turn to these carefully thought-out, theoretically driven coding schemes for a project outside the study of propaganda (and even there you may be hard pressed …). You are not likely to find them useful for a project investigating trends in gender or racial bias in advertisements, news, internet, or children's books.[25] Most current content analysis schemes are typically derived (inductively) in an interactive process between a careful reading of the text, design of preliminary coding categories, fitting of texts into these categories, and refinement of categories till most text can be fitted into the existing set of categories *given the specific research needs of the investigator.* The most typical reason given for the choice of a specific coding scheme is, indeed, its use in past research.[26]

Since content analysis deals mostly with written symbols and texts, early developers worked hard at carefully mapping their new technique into the different domains of linguistics: semantics, syntax, and pragmatics.[27]

They exclude pragmatics from the concerns of content analysis. "Content analysis restricts itself to semantics … With purely pragmatical characteristics … is not concerned."[28] Excluding pragmatics from the domain of content analysis leads early developers to restrict measurement to manifest content only; content analysis should not concern itself with latent content (intentional or unconscious meaning).[29] And since, in practice, content analysis has not been concerned with syntactical characteristics, although it could, in theory, deal with such characteristics, content analysis has basically restricted itself to semantics. Not surprisingly, one of the early collections on content analysis bears the title *Language of Politics: Studies in Quantitative Semantics* (Lasswell et al., 1949).

Within these texts, the early developers worked hard at specifying the amount of text within which to count the appearance of a reference, distinguishing between recording and context units.[30] And since content analysis produces numbers as frequencies, numbers that need to be analyzed statistically, they worked hard at pointing out the implications of using different units and different bases to deflate their frequencies (frequencies compared to what?) and at specifying strategies for the statistical treatment of results.[31] And since they were well aware that, back then as today, content analysis is a very labor-intensive process, and therefore expensive, they recommend sampling the amount of textual material to be content analyzed. The insights that Mintz has to offer on sampling[32] are truly precious.

While pushing hard for the development of the technical aspects of content analysis, Harold D. Lasswell and his associates displayed a remarkable awareness of its weak points. Harold D. Lasswell writes (*1942: 19): "Within any given set of symbol categories results depend upon technical coding operations. But the categories themselves exercise the most fundamental influence on results." As Berelson later put it, in an expression to become popular: "Content analysis stands or falls by its categories."[33] And so does the coders' interpretative process of fitting text (or symbols) into these categories. Content analysis "operations require judgments of *meanings* … Such judgments may be in error …" (Janis, 1949: 81). Not surprisingly, they pay close attention to issues of intercoder reliability as "an extremely serious [problem]."[34] "In what degree would other analysts arrive at the same results?" (*Kaplan and Goldsen, 1949: 84). To answer that question affirmatively, they recommend avoiding to code latent meanings ("It is an interpretation." Harold D. Lasswell writes, *1941: 2) and "intensive training of the analysts and … detailed rules of classification" as a way of increasing the reliability of content analysis categories.[35] They recommend always computing intercoder reliability for specific categories and groups of coders.[36] They recommend that "the degree of agreement among different coders in the application of the [coding] scheme,"

or intercoder reliability, should be "among the considerations which lead to the adoption or rejection of a proposed content analysis scheme."[37] They also recommend a design based on mutually exclusive categories, although there may be "prohibitive" "practical difficulties of achieving logical exhaustiveness" and only "empirical exhaustiveness" may be possible, perhaps "facilitated by constructing a list of which at least some subdivisions are logically exhaustive between themselves" (*Leites and de Sola Pool, 1942: 8). But even achieving that kind of exhaustiveness may be impractical and "only at a great expenditure of effort."[38]

The early developers' concern with technical issues did not entirely prevent them from tackling more general issues. Indeed, they attempted to set the technique in broader contexts, vis-à-vis other methodological approaches or general theoretical frameworks. Both Harold D. Lasswell's 1942 paper "Analyzing the Content of Mass Communication: A Brief Introduction" and Kaplan's 1943 "Content Analysis and the Theory of Signs" are good cases in point. Harold D. Lasswell's stated aim is to frame the technique in the context of a more general theory of politics. Kaplan similarly attempts to link quantitative content analysis not only to a more qualitative general theory of signs, but also to a general theory of politics and even to sociology since "content analysis is doubly sociological. … Not only are the propositions and terms of content analysis itself part of social science … but the propositions and terms *analyzed* are those which play certain roles in interpersonal relationships" (*Kaplan, 1943: 239; original emphasis).

And, yet, for all this emphasis on methodology, methodology of what? Data collection or data analysis? There is confusion on this issue. Viney (1986: 59), in her assessment of the use of the methodology in psychology, states: "Data collection and data analysis are two separate phases of the process of assessment. Content analysis of verbal communications is a form of data analysis. As such it is applicable to verbal data which have been collected in a variety of ways." For Altheide (*1987: 66), quantitative content analysis "provides a way of obtaining data to measure the frequency and variety of messages. … Data collection and organization (coding) …." Content analysis in this view is a method of data collection. In the end, which method is it?

No doubt, content analysis needs a set of texts as input, texts that must be *collected* according to some theoretical and statistical rules of selection (sampling). To the extent that data collection refers to what goes into content analysis as input, content analysis itself refers to the data analysis phase of those texts collected via other methods (archival work, in-depth interviews, psychoanalytic session, etc.). But if the purpose of content analysis is to supply numbers (frequencies of occurrences) to statistical procedures, then content analysis is simply a data collection method, no different from survey research (with which it shares basic similarities[39]).

But there is more to the confusion. In fact, the two views of content analysis as data collection and data analysis reflect perhaps different underlying epistemological views of the research process (see * Altheide, 1987, for an excellent rendering of the two approaches, Table 1, p. 67, in particular). Qualitative content analysis, as all qualitative approaches, does not draw a sharp separation between data collection and analysis; the two processes proceed in parallel and, simultaneously, in a reflexive interaction with the text. Quantitative content analysis, on the other hand, with its clear separation of the moment of coding scheme design (in the hands of the investigator) and coding (in the hands of the coders typically unaware of the broad theoretical aspects of a project), draws a sharp distinction between content analysis as a data collection technique, data to be analyzed at a later phase, after data collection, through statistical data analysis perhaps carried out by the investigator but certainly not by the coders ("novices hired and quickly 'trained'," *Altheide, 1987: 68).[40] Content analysis then *is* a method of data collection. But, perhaps more precisely, "content analysis [is] a technique of measurement applied to text" (where the coder is an instrument of measurement; *Markoff et al., 1975: 20, 35–38).

The leap in the methodological and theoretical sophistication of the contributions that grew out of Harold D. Lasswell's research project clearly stands out when we compare them with some of the best-known examples in this literature just a handful of years earlier, in the 1930s (grouped here under the heading "The Precursors"). Even Dale's article (*1932) dedicated to purely methodological issues on the analysis of content of motion pictures, is naive compared to the treatment of Harold D. Lasswell's research group. The coding scheme, which is not published, is ased on an acknowledged "common-sense classification … similar to that that lay adults commonly use for the description of motion pictures. … crime, sex, love, mystery, war, children, history, travel, comedy, and social propaganda. Subclassifications were drawn up under each of these categories …" (p. 246). Dale raises the question of intercoder reliability: "Does this method of classification yield uniform results when utilized by trained workers?" (p. 246). Needless to say, he does find a high degree of intercoder reliability (when reported, this is a standard finding in the literature!) and insists that several coders are necessary in order to minimize subjectivity. He confidently assures the reader that "An analysis of this type makes possible the answering of many important questions concerning motion-picture content" (p. 249).

Foster's article (*1935) on the change in stance of the United States from neutrality to involvement in World War I is based on a quantitative content analysis (*ante litteram*) of non-better specified "sample issues of the *New York Times* and of the daily press of Chicago" (p. 465), yielding some 11,000 separate news items.[41] The "five periods utilized," presumably between 1914 and 1917 (at least, according to the article title), "for

the purpose of detecting news trends which accompanied the evolution of American belligerence" (p. 467), are outlined in words but no definite beginning and ending dates for each period are given. The coding scheme ("a dozen attributes of each of these [11,000] news items" is not reported (in a bad practice that was to become standard in the literature, to this date). Issues of subjectivity and intercoder reliability – so central to the scholars involved in the *Experimental Division for the Study of War Time Communications* – are only superficially treated: "Many of the data were absolutely objective in character ... and the remainder were based upon classification so definite that the subjective element was reduced to the minimum" (p. 465). Findings of such complex research machinery –11,000 separate news items identified, a dozen attributes on each of these news items "recorded and transferred to punch-cards to permit mechanical compilation and to facilitate the making of correlations" (p. 465) – are presented as simple percentages here and there throughout the text (but, in any case, only a handful of them); there are no tables and no graphs, and not a single correlation.

Albig's study (*1938) of the program content of nine American and one English broadcasting radio stations from 1925 to 1935 reveals much greater attention to methodological issues. Beginning and ending dates for the sample of "four weeks out of each year for each station" are clearly indicated[42] and so are the names of the radio stations used (p. 343). The coding scheme is discussed (and published). Although Albig reassures the reader that "the categories for program classification were not arbitrarily devised," you will not find there any of Harold D. Lasswell's or Kaplan's awareness of the theoretical foundations of coding scheme design. Intercoder reliability is discussed and presented in the form of an overall correlation coefficient (p. 343). Findings are presented as percentages in three large tables that occupy the entire page; they are discussed seriatim, one coding category after the other, without any attempt to extract broader meanings or interrelations.[43]

Further Methodological Developments

The early developers truly defined the new technique of content analysis, more or less as we still know it today. They tackled all the issues that would later become staple chapters and sections in content analysis textbooks, from coding scheme design to intercoder reliability, measurement units, latent and manifest content, types of content analysis (qualitative vs. quantitative, thematic vs. referential), sampling, statistical treatment of results, data collection costs.

Of course, there have been refinements over the years. Sampling issues have received continued attention (e.g., Riffe et al., 1996). Statistical

measures of intercoder reliability have multiplied and their properties have been empirically investigated (e.g., *Krippendorff, 2004). Intercoder reliability has been extended to include the very selection of the material to be content analyzed. After all, different coders may have different ideas about what constitutes a "science article" or "sexual violence" (for these examples, see *Bauer et al., 2006: 105; *Malamuth and Spinner, 1980: 234). The separation between quantitative and qualitative content analysis has grown sharper. While Altheide (*1987) and others (e.g., Starosta, 1984, 1988) still use the label "content analysis" qualified as qualitative, very specialized approaches to the qualitative study of texts/symbols have now emerged (e.g., discourse analysis, conversational analysis, narrative analysis, semiotics).[44] In any case, in particular contexts, a quantitative analysis of texts based on frequency counts may be misleading. A police interrogator or a judge is interested in that single occurrence of such expressions as "I did it," "I murdered him," "I am guilty," in hours of interrogations or trial sessions. For these reasons, George (*1949) argues that, under certain circumstances, non-frequency content analysis provides more accurate information about sociohistorical reality than one based on frequencies.

The very definition of content analysis has also slightly changed with a greater emphasis on its inferential character. "Content analysis is any research technique for making inferences by systematically and objectively identifying specified characteristics within text" (Stone et al., 1966: 5). "Content analysis is a research technique for making replicable and valid inferences from data to their context" (Krippendorff, 1980: 21). In answer to the question "what's content analysis good for?" Roberts (1997b: 283) answers: "quantitative text analysis is good for drawing inferences about contextual and text-based variables …" (a goal underscored by the very subtitle of his book *Methods for Drawing Statistical Inferences from Texts and Transcripts*). Last but not least, seventy years on, the references to punch-machines ring quaint.[45] Even quainter is the reference to the "computor." Listen to how Harold D. Lasswell and associates (1942: 17; emphasis added) describe one of "the steps involved in applying any classification to newspapers and converting the results into statistically satisfactory form: … 4. the tabulation sheets then go to the *computor* [sic!], who notes the frequency of the indulgence and deprivation of each symbol in a table … and divides this frequency by the number of captions (or editorial inches) to get the final ratios." There was, indeed, a time when computers were human, mostly women, and that story is just starting to be told (e.g., Grier, 2005).

Computer-Based Content Analysis

Computer-aided content analysis was one of the most promising developments in content analysis in the 1960s, with the General Inquirer as the

best-known example. Developed by Philip Stone and his collaborators (*1966), it was heralded (by some) as "a system of computer programs to do content analysis objectively, automatically, and relatively painlessly" that has "given content analysis a new and respectable life" (Green, 1967: 397).

But how does the General Inquirer work? Basically, like any other content analysis application, you start from a coding scheme.[46] Beyond that, since it is the computer doing the coding, you need to supply it with as extensive a dictionary as possible, with words, expressions, etc., clearly assigned to a coding category.[47] The General Inquirer will search in a text for the symbols listed in the dictionary and tally their frequency of occurrence for each coding category. In the end, on these numbers, factor analysis can provide inferential material for substantive interpretations.[48]

By and large, the General Inquirer failed to deliver on early enthusiastic expectations. In Psathas's evaluation (1967: 174): "For those who wish to focus their research on computer program development for natural language processing, the General Inquirer already provides a reality; for those who wish substantive results by applying a 'ready-made' system, *caveat emptor.*" No doubt, reliance on the generic dictionary and on the coding categories supplied with the General Inquirer could only go so far in meeting the varied substantive needs of different investigators; and the alternative of producing ad-hoc coding schemes and dictionaries would certainly be very time consuming and expensive. Yet, if the proof of the pudding is in the eating, work based on the General Inquirer did produce interesting substantive results. Stone et al. (1966) provide several such substantive examples in their book. Namenwirth's (*1969) analysis of British newspaper editorials is another good example. Using 99 categories adapted from the Harvard General Inquirer Dictionary, several hundred most frequently used words (in the editorials), and factor analysis, Namenwirth analyzes a sample of 144 editorials that appeared in 1963 on Atlantic and European issues in three prestige and three mass British papers. The results highlight five major criteria of newsworthiness of events (*1969: 358) and a clear dividing line between elite and mass papers.[49]

In any case, the General Inquirer did represent a pioneering approach to automated content analysis. Much automated content analysis to this date is based on the General Inquirer approach to software design with its automatic matching of an input text to entries in an internal dictionary, classified into coding categories.[50] Such is the case of Gottschalk's approach to producing psychoanalytic anxiety scales (or other psychological states) from automated content analysis of transcript interviews.[51] Such is also the case of Laver's approach to the study of party policy positions (Laver, 1998; *Laver and Garry, 2000). According to Laver and Garry, computer-aided content analysis of party positions in the UK and Ireland in the 1992 and 1997 elections performed well against other common types of analysis (e.g., expert coding), but … at a much lower cost.[52]

Furthermore, once the dictionary has been properly "calibrated," it becomes possible "to computer code vast volumes of 'virgin' text in quantities that would be quite beyond the resources of any expert-coding technique" (*Laver and Garry, 2000: 633). And although computers run into problems in disambiguating words (taking words out of context, thus perhaps placing some in the wrong category), that problem of validity is more than offset by "a very significant gain in … *reliability*" (*Laver and Garry, 2000: 625).

Automated content analysis based on dictionaries, although perhaps less time consuming than expert coding, has the disadvantage that any meaningful application of the technique requires the time-consuming construction of special purpose dictionaries (for this critique, see Psathas, 1967). Artificial Intelligence (AI) approaches to computer understanding of natural languages try to avoid this problem through an automated parser that "receives input text … and breaks it up into parts of speech like nouns, verbs, and attributes … draw[ing] upon both syntactical rules and semantic relations to assign meanings to classes of words" (*Bond et al., 2003: 737).[53] AI automatic parsing approaches are still far from delivering valid data (a historian would find particularly troublesome the duplication of historical events …), but if gross patterns are all that is needed (or first approximations for more in-depth and more detailed studies), then these approaches will serve well. The same can be said about purely statistical approaches to text, in particular those born out of the French school of *analyse des données textuelles* involving factor analysis of correspondence (*Guerin-Pace, 1998: 76).[54] Indeed, as Guerin-Pace writes (*1998: 73): "The use of statistical methods of textual analysis [in the form of factor analysis and graphical representations] offers an extremely rich exploratory approach …" In Tukey's words (1977: 3): "Exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone – as the first step."

Linguistics-Based Content Analysis

In his 1938 article "A Provisional Classification of Symbol Data" (in the journal *Psychiatry*), Harold D. Lasswell proposes an interesting coding scheme, centered around such actors as the self and nonself (the nonself as the other, including the psychiatrist interviewer, or as a thing), their interrelationships, and the nature of these interrelationships (positive or negative, pro or anti the other in any dyad). What is interesting about this coding scheme[55] is the idea of a basic template actor–action–actor, or subject–action–object, but where neither actors nor actions are precisely measured (e.g., the actor as "mother" or "boyfriend" and the action as "kissed" or "died").

Gottschalk, in his computer-assisted applications of content analysis in psychoanalytical research, would take Harold D. Lasswell's insight onboard, when he complains that computerized single-word or single-phrase approaches "fail to identify who did or felt what about whom."[56] But it is Hall and de Castle's study of dreams (1966) that takes that subject–action–object approach to the limits. There are characters (or actors), with given characteristics; there are their activities in time and space; and there are interactions (e.g., aggression, friendliness, love/sex) and emotions (e.g., anger, apprehension, happiness). Considering that these were the days when information coded in a coding scheme would be punched in an 80-column card where space was at a premium (and therefore all codes would be abbreviated to one- or two-letter symbols), these applications of content analysis are remarkable.

Further developments along the actor–action–actor template (or subject–action–object, SAO), however, would require two things: one, the application of linguistics, theories of narrative in particular, in content analysis; and two, increased computer capacity. Voices recommending to bring linguistics into content analysis had been heard over the years (on these issues, see Franzosi, 2004a: 40–41; see also *Saporta and Sebeok, 1959; *Hays, 1969; *Markoff et al., 1975: 28–31). Linguistics, though, only served the purpose of providing *caveats* against facile views of objectivity of coding, of meaning being captured by a codebook or by training. The first concrete examples of coding schemes based on a body of linguistic theory were late to come. But when they came – from scholars with different disciplinary backgrounds, different purposes, and different methodological and theoretical orientations, conscious or unconscious of the linguistics underpinnings of their work (or just paying lip-service to it) – they all shared remarkably similar characteristics, centered around the basic structure SAO (or SVO, subject–action–object, or subject–verb–object), the "canonical form" of language or noun phrase–verb phrase (see *Katz et al., 1969; Markoff et al., 1975; *Franzosi, 1989; *Carley, 1993; *Abell, 2004; see also Abell, 1987; Heise, 1989; Corsaro and Heise, 1990).

There are good reasons for this. After all, social action is about actors – individuals, groups, organizations – doing or saying something, pro or against someone else. Narrative renders social action/social relations at the linguistic level. There is a homologous relationship between narrative and social action/social relations (Franzosi, 2004a: 297). But narrative naturally translates into the SAO structure (Franzosi, 1989, 2004a: 298–299). To the extent that social scientists are interested in measuring social action, they will encounter the SAO structure, by hook or by crook, whether they know any linguistics or not (see Franzosi, 2004a: 43–51, on the development of coding schemes based on the SAO structure). And this narrative understanding and measurement of social reality shifts social scientists'

focus of explanation from variables to narrative (see *Abell, 2004; Franzosi, 2004a: 238–246). The application of such complex linguistics-based coding schemes to sociohistorical questions involving large-scale data collection projects shows that these schemes work well beyond toy examples (see here *Markoff, 1990; *Franzosi, 1999).

Texts (narrative texts in particular) will not just link social actors across given spheres of action (e.g., communication, conflict, violence, emotions, linking one actor to another). Regardless of text genre, texts will link concepts, thoughts, images, ideas. Literary critic Kenneth Burke long since expressed that basic idea (1957: 18):

> The work of every writer contains a set of implicit equations. He uses "associational clusters." And you may, by examining his work, find "what goes with what" in these clusters – what kind of acts and images and personalities go with his notions of heroism, villainy, consolation, despair, etc. And though he may be perfectly conscious of the act of writing, conscious of selecting a certain kind of imagery to reinforce a certain kind of mood, etc., he cannot possibly be conscious of the interrelationships among all these equations. Afterwards, my inspecting his work "statistically," we or he may disclose by objective citation the structure of motivation operating here. There is no need to "supply" motives. The interrelationships themselves are his *motives.* For they are his *situation*; and *situation* is but another word for *motives.* The motivation out of which he writes is synonymous of the structural way in which he puts events and values together when he writes; and however consciously he may go about such work, there is a kind of generalization about these interrelations that he could not have been conscious of, since the generalization could be made by the kind of inspection that is possible only *after the completion* of the work.

It is to these types of structures that Carley (*1993) and Bearman and Stovel (*2000) dedicate their attention, indeed trying to uncover them "statistically," through map analysis and network models.

What Next?

Some twenty years ago, Markoff et al. (1975: 9) wrote: "debates over the *value* of content analysis, however frequent, are foolish; for if the technique pervades the social studies, our only reasonable policy is not to approve it or disapprove it. But to improve it." And so it was. The trickle of methodological innovations traced here – as much the result of technological advances in computer hardware and software as of theoretical work – makes that very clear. And if it had been Altheide's vision of an

ethnographic "content analysis" that brought together a traditional quantitative thematic approach to text to a more qualitative, "narrative" approach, recent developments reviewed here show that even narrative can now be quantified, bringing together classical thematic content analysis with narrative analysis under a single quantitative framework.

And yet, problems remain. Seventy years on and content analysis is still an expensive research tool. And it is so even in computer-assisted content analysis. Whether manually entering the information from an external medium (e.g., Franzosi's PC-ACE, Program for Computer Assisted Coding of Events, www.pc-ace.com) or tagging textual expressions before files are read into a software (e.g., Carley's Automat, or any of the standard approaches by qualitative data analysis software, such as Maxqda, N6, or Atlas.ti), computer-aided content analysis is *still* time consuming. Paradoxically, content analysis software has allowed investigators to devise far more complex coding schemes (e.g., story grammars based on pages of rewrite rules) that, while yielding far richer data, also require greater coding time. Not surprisingly, investigators have either resorted to automatic statistical analyses on input text (e.g., *Guerin-Pace, 1998) or attempted to automate the coding process, eliminating the "unreliable" human coder (e.g., *Laver and Garry, 2000; *Bond et al., 2003). The frontier of content analysis, no doubt, lies here. Computer understanding of natural language is no longer an esoteric pursuit for a handful of interested academics. The huge amounts of symbolic data available on the web make it a commercial imperative. A solution will be found. It is only a question of time.

But while waiting for that time, however short, there is yet another frontier that content analysis practitioners should push (after that of linguistics): the application of rhetoric in the construction of coding schemes.[57] It is truly remarkable that even those applications of content analysis where knowledge of rhetoric would be most helpful (e.g., *Katz et al.'s study of persuasive appeals, 1969 or *Gamson and Modigliani's study of "frames," 1989) make no reference to rhetoric. Yet, through the centuries, rhetoric has built a large *taxonomy of what to look for in a text.* And for centuries, school and university education was based on the *trivium*: logic, grammar, and *rhetoric.* In antebellum American South, "male academies and colleges" taught rhetoric in preparation to a man's life in politics or the professions (where the art of persuasion was crucial). Since Southern "women did not make speeches," the first women's colleges substituted the study of rhetoric with that of botany, another discipline also "largely concerned with taxonomy" (Farnham, 1994: 75, 81).[58] And if botanic taxonomy is of little help to improve the methodology of content analysis, rhetorical figures should provide indispensable tools of textual analysis, particularly when investigators' interest lies in forms of argumentation.

Content Analysis on What?

Back in 1943, in the take-off days of content analysis, Kaplan (1943: 237, 239) wrote: "content analysis, as at present developed, is directed toward the analysis of political discourse, not of discourse in general, nor of the various types of discourse … with propaganda, public opinion, ideologies, and related matters." The contributions collected here show the far wider field of application of content analysis beyond Kaplan's restricted scope. From politics to culture, from gender to race, from rap music to advertising, from mental health and psychiatric disorders to terror and violence, from the internet to television, radio, newspapers, magazines, books, party platforms, journal articles, and interview transcripts, from text to photographs, cartoons, websites, films, and videos, the range of topics, media, and symbolic material analyzed has grown exponentially over time.[59]

No doubt, the technique of content analysis has seen a surge in popularity in recent decades. Riffe and Freitag (*1997), in their content analysis of 486 articles published in the *Journalism and Mass Communication Quarterly* between 1971 and 1995, found evidence of a positive trend in the number of research articles based on the technique.[60] Newspapers (46.7% followed by TV at 24.3%) and news (71%, followed by advertising at 10%) are the most popular medium and content focus. Worrisomely, they found that issues of intercoder reliability are not addressed with due care (only 56% report intercoder reliability coefficients, typically overall rather than by individual coding category and in any case only 10% on a random sample of coded content).[61]

Interestingly, Riffe and Freitag did not feel the necessity to record whether coding schemes were reported in the articles analyzed – yet, any replication of a content analysis study would require access to the coding scheme used in the study (and perhaps even to coding instructions). Nothing new in this: few of the application studies collected in these volumes report the coding scheme (and certainly, none report the code book of definitions of each coding category). There are certainly good editorial reasons for this practice. The publishing of the coding scheme (not to mention of the code book) would add far too many pages to the normal length of a journal article. Yet, this practice has meant that every scholar, seasoned or novice, using content analysis as the main methodological research tool, has had to start from scratch in devising content analysis schemes and that replications of studies are nearly impossible. Listen to what two of the founders of the technique, Leites and de Sola Pool, wrote back in 1942 (1942: 14):

It is essential that the report of results include as full a reproduction of the definitions of the categories employed as possible. Only then will it be possible for other teams or individual investigators to

> relate their own work to that of the team involved. Here as else-
> where the rush for the communication of results at the expense of
> the presentation of methods is scientifically self-defeating.

An early lesson that unfortunately has fallen on deaf ears, no doubt con-
tributing to a widespread general negative feeling about content analysis
as a somewhat "suspect"[62] and ad-hoc technique.

Riffe and Freitag also found that content analysis articles are typically
descriptive in nature: articles are not grounded in an explicit theoretical
framework (only 27.6% of them), they do not have explicit research ques-
tions or test hypotheses (only 20% do), and they rely on simple, descrip-
tive statistics (40.1%). Nothing new here. As Albig (*1938: 349) had
written decades earlier: "the most valuable use of studies of content … is
in noting trends and changes in content." Indeed, for all the emphasis of
early developers on hypothesis testing, most of the applications in this
collection have a common concern with a descriptive mapping of "trends
and changes in content," whether over time or cross-nationally, whether
looking at newspapers, television, or the internet, advertising or news,
whether about gender, race, or politics. Even when research questions are
explicitly stated, these questions are typically descriptive, as exemplified
by: "Is coverage of specific space elements concentrated in specific sec-
tions of the *New York Times?*" (*Clark and Illman, 2003: 24), "how often
[do] online newspapers use hyperlinks"? (*Dimitrova et al., 2003: 407), or
"what proportion of disaster relief home pages have high, moderate, and
low levels of interactivity?" (*Paul, 2001: 743).[63]

Yet, there is description and description. For one thing, the best studies
will not just report percentages one category at a time. They will correlate
the results across categories, asking, for instance, how different types of
media (e.g., sensationalist vs. serious news outlets) frame news on political
issues (*Semetko and Valkenburg, 2000) or how sex roles in advertise-
ments change by product category or time period (*Ferguson et al., 1990).
Furthermore, even when used descriptively, content analysis does not
necessarily just deliver "trends and changes in content." It can highlight
the mechanisms behind those trends and changes. The openly derogatory
portrayal of American blacks has declined; yet, content analysis reveals
how blacks and whites, when portrayed together, do not really interact
(e.g., *Humphrey and Schuman, 1984: 560–562; *Pescosolido et al., 1997:
455). Races in America continue to be "separate and unequal."[64]
Similarly, the role of women in society has changed, and their represen-
tation has partly reflected those changes; yet, women are more likely to be
portrayed as young, attractive, in non work-related settings, and in allur-
ing images (*Gilly, 1988; *Ferguson et al., 1990: 46–47).

It is, perhaps, in light of the descriptive nature of most content analysis
applications that Ferguson et al. (*1990: 49) concluded: "content analysis is

inherently descriptive and can provide only limited insight into *why* signif-
icant relationships or trends are observed." Yet, is the technique *inherently*
descriptive? True. The word "description" features prominently in
Berelson's definition of content analysis. Yet, content analysis *can* be used
to address the *why* question. But that would require additional information
(or data on independent or explanatory variables, if you wish), information
necessary in order to explain the "trends and changes in content" provided
by content analysis. Thus, Griswold convincingly shows that US copyright
laws (and not the lack of American writers' interest in themes of love and
marriage popular among their European colleagues) were responsible for
the "unique set of themes and subjects" found in nineteenth century
American novels (*1981: 747). Mueller (*1997) uses protest event counts
collected via content analysis from various newspapers to test Snyder and
Kelly's model of media sensitivity and event intensity in the enumeration
of protest events. Pescosolido et al. (*1997), in their study of the portrayal
of blacks in US children's picture books, provide ample evidence on the
silence of blacks in children's books; but they also tackle the *why* question
(*1997: 456–458): was there any relationship between the greater visibil-
ity of blacks in children's books after the late 1960s and a surge in racial
conflict during that same decade? Jackson et al. (* 2007) similarly use con-
tent analysis of job advertisements in local and national British newspapers
to provide data that can explain the declining importance of education in
class mobility.[65] Whether used in the context of hypothesis testing, when
content analysis data are integrated with other kinds of data (in particular,
experimental design or survey data) greater insights can be gained (e.g.,
*Min, 2002; * Johnson et al., 1971; *Gamson and Modigliani, 1989).

Not that work based on hypothesis testing is necessarily "better" and
descriptive work "worse." As I have written: "The idol of 'hypothesis test-
ing' before which we kneel with religious fervor tends to obscure the
invaluable role that good statistical work can play in bringing out patterns
in the data" (Franzosi, 2004a: 231). Indeed, a ritualistic use of hypothesis
testing and multivariate work may end up hiding more than it reveals.
After all, as Lieberson wrote (1985: 213): "one of the really valuable func-
tions of empirical social research is a descriptive one." William Whewell,
in his massive work on the history of science (1837, 1840) became a fer-
vent advocate of the Baconian inductive method based on the careful "col-
ligation of facts." The sad and recurrent fact ("consistent across different
countries during these 25 years [1975–1997]"; *Furnham and Mak, 1999:
431) that, decades after the feminist movement, women are *still* portrayed
in the media stereotypically, negatively, or at least unequally to men (and
perhaps in ways more subtle and harder to detect) – paradoxically, even
in magazines, such as *Ms.*, "with an *explicit commitment* to nonstereotypical
portrayal of women" (*Ferguson et al., 1990: 48) – is an important finding,
even if it addresses only the *how*, rather than the *why* question (*Gilly, 1988;

*Lindner, 2004; *Malamuth and Spinner, 1980). And there is no comfort in knowing that the aesthetics of the male body is starting to be subjected to the same kind of impossible standards of attainment as the wafer female body (* Law and Peixoto Labre, 2002: 706). The same is true for American blacks, decades after the Civil Rights Movement. While the most degrading portrayals of American blacks, once typical (*Johnson and Sears, 1971: 710), have disappeared, the portrayal is still far from equal vis-à-vis whites, whether one looks at news, advertisements, or children's books (*Johnson and Sears, 1971; *Clawson and Trice, 2000; *Humphrey and Schuman, 1984; *Pescosolido et al., 1997).

Knowing *how* different interpretative media representations ("packages") of nuclear power have fared over time (*Gamson and Modigliani, 1989), *how* US public service announcements for television broadcast HIV/AIDS prevention rhetorically construct the risk's of contagion (*Dejong et al., 2001), *how* the design characteristics of advertisements affect memory recall (*Naccarato and Neuendorf, 1998), *how* the European press and television news differently frame European politics (*Semetko and Valkenburg, 2000), *how* US news websites frame the Iraq War and act as information gatekeepers (*Schwalbe, 2006; see also *King and Lester, 2005, on newspaper representations), *how* a shift toward a more conservative stance was marked by successive US presidents verbally in their public speeches (*Prothro, 1956), *how* the internet provides a (limited) interactive medium in disaster situations (*Paul, 2001) or opportunities for self-presentation in personal home pages (*Dominick, 1999), *how* the European Women's Lobby website frames its discourse (e.g., the scant use of the word *feminist*, or its narrow focus on Europe; *Pudrovska and Marx Ferree, 2004), *how* music videos provide somewhat expected and stereotypical images of race and gender (* Brown and Campbell, 1986), and *how* heavy metal and rap music are perceived in the media (*Binder, 1993), *how* news concerning space exploration and exploitation is presented (*Clark and Illman, 2003), or *how* science is reported and used in different institutional contexts (democratic UK or communist Bulgaria) (*Bauer et al., 2006) *is* valuable knowledge. And that science needs ascertained facts and reliable data patterns as much as theories (or, more simply, "generalizable description of communications content"; Kassarjian, 1977: 10).[66]

Conclusions

Nearly seventy years have passed since the first recorded use of the label "content analysis." From its original focus on propaganda and politics, the label has come to encompass other (and even earlier) approaches to the study of content (content of the press, editorial content, content of

mass communication, … .), all joined together by a reliance on a coding
scheme based on a set of coding categories, a coder, and a body of text
(or more generally, of symbolic material, e.g., pictures) to which the
coder applies the coding scheme with the goal to quantify the frequency
of occurrences of coding categories. The parallels with a technique being
developed by Paul Lazarsfeld and others at around the same time – sur-
vey research – are striking: the questionnaire, the interviewer, and a body
of people (sample) to whom the questionnaire is distributed for the pur-
pose of obtaining frequencies of responses to questionnaire items.[67] No
doubt, those were the days of a firm belief in the power of quantification.
As Pitirim Sorokin (1933: 196) put it: "In the future some thoughtful
investigator will probably write a very illuminating study about the
'quantitative obsession' of a great many social scientists, psychologists
and educators of the first third of the twentieth century, tell how such a
belief became a vogue, how social investigators tried to 'measure' every-
thing." These four volumes tell the story of one way in which social inves-
tigators tried to measure everything: the content of text (or, more generally,
of symbolic material).

Substantively, the applications collected here leave no doubt about
the widespread use of the technique across different fields (media, adver-
tising, communication, marketing, psychology, psychoanalysis, educa-
tion, political science, sociology, etc.). They certainly do not support
Janowitz's conclusion (1969: 653) that "there is a kind of intellectual stag-
nation in the field [of content analysis]. The standard methodological
treatise on content analysis, prepared by Bernard Berelson, went out of
print ten years after its original publication. It has not been revised, nor
are there signs of important new efforts in this direction" – untimely con-
clusion, in light of the appearance of two new textbooks in those years
and of the proceedings of the conference on content analysis which had
gathered in 1967 over 400 participants at the Annenberg School of
Communication[68] (Budd et al., 1967; Holsti, 1969; Gerbner et al., 1969).

In the end … , here are my recommendations, if these pages and this
collection of papers have inspired you to try your hand at the method-
ology of content analysis:

1. Try your best to be rigorous (in sampling input documents, in design-
   ing your coding scheme, in testing hypotheses). An honest attempt at
   rigor is better than fast and sloppy work and certainly better than a
   cheap, anything-flies approach to the research enterprise.
2. Pursue rigor, but don't fall prey to rigidity or to an unreflexive rhet-
   oric of objectivity and science (Franzosi, 2004a: 229–232). Provide
   stringent definitions of coding categories, train your coders, compute
   coefficients of intercoder reliability. Engage in "science." But don't fool

yourself into thinking that this will have solved the real issues of meas-
uring meaning. Several of the applications presented here[69] reveal
how the most intractable problems of meaning and interpretation are
really solved in the practice of content analysis: by consensus, through
team discussions (would a different team reach the same consensual
answer?).[70]

3. Be aware of the limitations of the technique. But don't put off tackling
   questions that may otherwise remain without answers. "In the ex-
   ploration of interesting questions, sometimes less than perfect techniques
   must be tolerated" (*Griswold, 1981: 763).

4. Do use crude statistical frequency counts of words or automatic com-
   puter- aided systems. The idea that you should not rely on automated
   systems because of their lack of validity. Indeed, they may serve you
   well for exploratory purposes. If lucky, even a "fishing expedition" can
   provide insights for further research.

5. Whenever you can, try to mix the quantitative and the qualitative.
   Never forget that, in content analysis as in survey research, your num-
   bers were once words, words that you turned into numbers by count-
   ing (Franzosi, 2004a: 287–297). Snippets of words will provide
   concrete examples of what's behind the numbers, so contextualize
   them, give them life (for a good example, see *Gamson and Modigliani,
   1989; see also their argument in using qualitative and quantitative
   data, 1989: 11).

6. Taking Altheide's advice one step further, don't just stop at the sound
   bytes in mixing quantitative and qualitative (Altheide's "narrative").
   You can even quantify narrative! To the extent possible, try to mix a
   quantitative thematic analysis and a quantitative narrative analysis.
   Most texts will not only have themes, but also (at least some) actors
   and actions. A thematic analysis would not give you the matrix of
   interaction among social actors around specific spheres of action and
   you would miss out.

7. "What goes with what" is relevant not just in terms of relations among
   social actors, along a basic SAO structure. It is also relevant in terms
   of the conceptual matrix of interaction of ideas, images, thoughts,
   metaphors that make up a text. If you focus on themes, don't just look
   at them *seriatim* but in their complex interrelations (what goes with
   what ; *Carley, 1993).

8. Familiarize yourself with the main rhetorical figures and bring rhetoric's
   taxonomy to bear in the construction of coding schemes. It will be well
   worth the investment!

That said … do enjoy this collection and, above all, enjoy your new
research project based on content analysis!

Notes

1. Holsti (1969: 2–3), Krippendorff (1980: 21–24), Shapiro and Markoff (1997: 11–13), Franzosi (2004a: 32–33).

2. For these definitions, see Janis and Fadner (1942: 2; emphasis added), Kaplan (*1943: 230, 243; emphasis added), Kaplan and Goldsen (*1949: 83; emphasis added). For a quick summary of definitions, see Holsti (1969: 2–3), Shapiro and Markoff (1997).

3. For the quotations in this paragraph, Leites and de Sola Pool (*1942: 1; emphasis added), Mintz (*1949[1943]: 127), Lasswell (*1942: 16; emphasis added), see also Janis (1943: 429).

4. "A social science sentence may be called one of 'content analysis' if … it indicate[s] frequencies of occurrence … it must assign numerical values to such frequencies." "The customary procedure in content analysis is to classify the material investigated … into a number of categories and to count their number …" Within a "body of content … the appearance of a reference is counted."

5. de Sola Pool (1959: 196) continues: "The units could vary greatly: there were counts of column inches, of key words, of themes, of literary forms, of types of characters, etc. … almost all studies had for their basic logic a comparison of the frequency of certain types of symbolic expression in different segments of text …" In one of the earliest statements by Harold D. Lasswell (1938: 198), before the label "content analysis," he wrote: "We … classify references into categories … References … may be quantified by counting the number of references which fall into each category …"

6. For the quotes in this paragraph, see Kaplan (*1943: 244; emphasis added), Janis and Fadner (1942: 1–2; emphasis added), Leites and de Sola Pool (*1942: 2), Berelson (1952: 18; emphasis added). For similar wording on impressionism, see also Kaplan and Goldsen (*1949: 83), Berelson (1952: 119).

7. Kaplan (*1943: 246; emphasis added). On hypothesis testing, see also Lasswell (1938; 1942: 3–4).

8. Even Waples and Berelson (1941: 1) in a note in their paper "What the Voters Were Told (An Essay in Content Analysis)" write: "The authors are heavily indebted to … Harold D. Lasswell, the virtual originator of objective methods of content analysis."

9. Willey (1926) talks about "newspaper analysis" (or "press analysis") and so does Woodward (1930: 39–64; 1934).

10. For a brief introduction to the history of the label, see also Franzosi (2004a: 32–35).

11. For notes on the biographies of Lasswell, Waples, and Berelson, the early developers of content analysis, and their scientific approach to social research, see Sills (1981), Berelson (1979), Richardson (1980), Rosten (1969). On Lasswell and Chicago sociology and political science, see also Sylvan (1991).

12. "Content analysis" appears several times in Chapter IV ("The Content of Publications") and provides the title to Appendix B ("Notes on Content Analysis").

13. Lasswell, in the early stages of development of the technique, was using such labels as "symbol analysis," "content of mass communication," "content of the press," "editorial content" (Lasswell, *1941, *1942; Lasswell and Associates, 1942; Lasswell, Geller, and Kaplan, 1942; see the volume by Smith, Lasswell, and Casey, 1946, for a good compendium of annotated bibliographies).

14. The Experimental Division had been set up at the beginning of the war in the basement of the Library of Congress in Washington, DC. The Division was part of the War Communications Research Project sponsored by the Rockefeller Foundation

INTRODUCTION   xliii

(*Lasswell, 1941; Lasswell, Leites, and Associates, 1949: v; Berelson, 1952: 23). Harold D. Lasswell, then a professor of political science at the University of Chicago, became its director.

15. The book brought together the proceedings of a 1955 conference in Monticello, Illinois, by the Committee on Linguistics and Psychology of the Social Science Research Council (de Sola Pool, 1959: 1).

16. See Budd et al. (1967), Holsti (1969), Krippendorff (1980), Weber (1990), Neuendorf (2002). See also Rosengren (1981). Yet, the proceedings from the 1967 conference held at the Annenberg School of Communication in Philadelphia were still published with the title *The Analysis of Communication Content* (Gerbner et al., 1969) - although the editors leave no doubt that the book is about content analysis, as they make clear in very "Preface."

17. Finding those documents, even in an age of computerized inter-library loans, has not been easy and for some of the internal documents of the *Experimental Division for the Study of War Time Communications* I had to recur to the revised versions published in Lasswell et al. (1949).

18. "Content analysis must give primary attention to … the selection of categories" (Waples et al., 1940: 149). Nearly the entire Appendix B of Waples et al.'s (1940) "Notes on Content Analysis" deals with coding scheme design. "See also Lasswell (*1942), Leites and de Sola Pool (*1942), Kaplan (*1943), Kaplan and Goldsen (*1949).

19. Lasswell (*1942: 1).

20. Even when Lasswell recommended content analysis in the study of problems lying outside propaganda and politics (e.g., psychoanalytic interview situations), he insisted that "One sound rule of choice [in coding categories] is to classify with a view of obtaining data which are relevant to definite hypotheses …" (Lasswell, 1938: 198).

21. Statements of demand make certain claims ("We want higher wages"); identification statements are about self-identity ("I am an Italian"); and fact statements are about given facts ("Violence erupted in the evening in the outskirts of Paris").

22. Lasswell also refers to positive and negative direction as indulgence and deprivation (*1942: 2). Furthermore, he proposes to measure the "intensity" or "magnitude" of the direction (*Lasswell, 1941: 3–4; *Lasswell, 1942: 8). On these definitions, see Lasswell (1941: 3–4; *1942: 27–29), Kaplan (*1943: 242), Kaplan and Goldsen (*1949: 84).

23. Kaplan and Goldsen (*1949: 84), Kaplan (*1943: 242), Lasswell and Associates (1942: 16).

24. Lasswell (*1942: 6), Kaplan (*1943: 242). Not only is the list potentially infinite, but "there is no cut-and-dried list of political symbols (and objects) that will serve the needs of every research on politically significant contents of the press" (Lasswell and Associates, 1942: 13).

25. Yet, Kaplan turns these limitations into strengths. True. A content analysis thus restricted works well with political discourse only and other types of discourse, scientific or esthetic/poetic, would require if not altogether different methods, certainly different categories. The categories elaborated for political discourse "obviously … would not be fruitful for the analysis of poetry." "It is obvious that for the analysis of scientific discourse or of poetry quite different basic categories would be desirable." But, on the positive side, "the relation between these categories and the political orientation is fairly direct;" there is a "direct connection of these categories of key symbols with the political process" and that connection "is clear … from the categories themselves." "The usefulness of these categories for the specific purposes of content

analysis is indicated by the readiness with which they suggest and the ease with which they adapt themselves to hypotheses [testing of political processes]." For the quotes in this paragraph, see Kaplan (*1943: 243), Kaplan (*1943: 241; see also pp. 237, 239, 242), Kaplan (*1943: 241, 242, 243), Kaplan (*1943: 241). I should point out that I use a similar argument to Kaplan's to defend the limits of Quantitative Narrative Analysis (QNA): that while QNA applies *only* to narrative texts, these types of texts map well on the coding schemes used in QNA, namely, story grammars (Franzosi, *1989: 295; 2004b: 555).

26. For acknowledged examples of this practice, see here, Clawson and Trice (*2000: 55), Tak et al. (*1997: 422), Gilly (*1988: 78), Ringold and Calfee (*1989: 8), Ferguson et al. (*1990: 43), Min (*2002: 930), Lindner (*2004: 413–414), Brown and Campbell (*1986: 99); on an inductive approach, see here Schwalbe (*2006: 270) and Pudrovska and Marx Ferree (*2004: 126) who acknowledge: "Our approach in delineating the set of categories was to make them emerge from the data."

27. See Lasswell (*1941: 2, 14), Kaplan (*1943: 234, 237), Leites and de Sola Pool (*1942: 6), Janis (1943: 432).

28. Kaplan (*1943: 234, 237). See also Leites and de Sola Pool (*1942: 6), Janis (1943: 432, 1949: 59).

29 See Lasswell (*1941: 2), Kaplan (*1943: 234, 237), Leites and de Sola Pool (*1942: 6), Janis (1943: 434).

30. In Lasswell's definition (*1942: 16), the recording unit is the smallest body of text within which to search for the occurrence of a reference; the context unit is the largest body of text within which a recording unit can occur (e.g., a sentence as the recording unit with a paragraph as the context unit). On the size of text units, see the discussion of Leites and de Sola Pool (*1942: 14–16).

31. On these three separate issues, see, respectively, Geller et al. (1942), Lasswell (*1942: 17–19), Lasswell and Associates (*1942: 13–15), and Leites and de Sola Pool (*1942: 18).

32. For example, Mintz's caveats on sampling when investigators are interested in changes in trend.

33. Berelson (1952: 147). *Stone et al. (1966: 134) would fully embrace that statement.

34. Janis (1949: 81). Leites and de Sola Pool (*1942: 12) refer to intercoder reliability as intersubjective stability of a coding scheme.

35. See Kaplan and Goldsen (*1949: 94–96), see also Janis (1949: 81). That recommendation is still embraced today (Roberts, 1997a: 110–111). Indeed, the recommendation of relying on a handbook of written instructions, precise definitions of categories, and training is, by and large, a sound one. Yet, the recommendation is likely to ignore the extensive background knowledge essential in any coding process (on these issues, see Franzosi, 2004a: 219–226; on the limited role of rules and training, see also *Ahuvia, 2001: 143–144). Proof be the findings by Peter and Lauf (*2002) that coders with greater political knowledge produce more reliable code in cross-national reliability studies of television news reporting of European politics (1999 European election campaign and routine coverage of the EU). Furthermore, as Hruschka et al. argue (*2004: 321–322), codebooks and training may simply lead to "interpretive convergence" among the members of a coding team, and specific to that team.

36. See Lasswell (*1942: 15), Janis (1949: 3).

37. See Leites and de Sola Pool (*1942: 12), see also Janis (1949: 81).

38. Leites and de Sola Pool (*1942: 11). Yet, that recommendation of exhaustiveness and mutual exclusivity of coding categories still features prominently among Holsti's "general requirements" of coding scheme design (1969: 95, 99–100).

39. On the similarities, see below; see also Franzosi (2004a: 4, 2004b: 561).

40. Interestingly enough, to add to the confusion, it is Viney, a quantitative content analyst, to express the view of content analysis as a data analysis methodology (contrary to expectations).

41. Presumably, more information on the sample of the NYT issues and on the specific Chicago newspapers used is reported in an unpublished document cited in a footnote.

42. The selection of these weeks is justified "as adequate" on the basis of non-better explained and not reported "sampling tests." (*Albig, 1938: 343).

43. In that sense, de Sola Pool (1959: 196) is right on target when he writes that by the late 1950s there emerged "a new kind of quantitative [content] analysis … more interested in contingencies than in straight frequencies. … Contingency analysis is a quantitative procedure. It involves counting. But the form of the hypothesis and of the critical observations is different from that in a simple frequency analysis. Contingency analysis asks not how often a given symbolic form appears in each of several bodies of text, but how often it appears in conjunction with other symbolic units."

44. See, for all, the volumes in this Sage series: van Dijk (2007), Drew (2006), Atkinson (2006), Gottdiener (2003).

45. See, for instance, Foster (*1935: 465).

46. Stone et al. (1966: 134) repeat Berelson's statement (1952: 147) that a coding scheme/dictionary is vital for successful automated content analysis(*1966:134); see also Laver and Garry (*2000: 626). For computer-aided word-count content analysis based on large dictionaries, Weber (*1983: 143) argues that "for general-purpose dictionaries the content classification scheme has little or no effect on the substantive results."

47. Several different dictionaries were produced, the best known being the Lasswell Value Dictionary and the Harvard dictionary (see *Weber, F. 1983: 127, 132). On early available dictionaries for the General Inquirer, see Stone et al. (*1966: 140–141).

48. Factor analysis, or similar techniques, in content analysis is still a popular technique of data analysis (e.g., *Semetko and Valkenburg, 2000, *Cole, 2005, *Jackson et al., 2005).

49. Namenwirth's (1973) content analysis of American Republican and Democratic party platforms from 1844 through 1964 was similarly based on the General Inquirer (73 categories where "95 percent of the words which occur in party platforms were entered in a dictionary") (p. 650). Namenwirth argues that the history of value change is not linear but cyclical. Namenwirth's study became part of a larger study by Namenwirth and Weber (1987) on the temporal dynamic of cultural indicators, also based on the General Inquirer. Namenwirth and Weber, besides American party platforms, analyzed speeches from the British throne at the opening of Parliament, and presidential speeches from three scientific societies.

50. For a brief overview of computer approaches to content analysis, see Stone (1997).

51. See Gottschalk and Bechtel (*1995), see also Gottschalk and Gleser (1969), Gottschalk (1979), Gottschalk et al. (1986).

52. On the relative performance of different approaches to inferring party ideological positions, see Gabel and Huber (*2000).

53. For automatic coding of event-types data, see also the KEDS projects (Kansas Event Data System) (Schrodt, 2006).

54. See also Lebart (1993), Lebart and Salem (1994), Lebart et al. (1998), Bolasco (2005). See Guerin-Pace (1998: 76) on software options; see also Bolasco's TALTAC at http://www.taltac.it/it/index.shtml.

55. This type of coding scheme would later appear in Lasswell's basic coding scheme for the study of propaganda and political communication and not just dreams or psychoanalytical sessions (Lasswell, *1941: 11, *1942: 25).

56. Gottschalk and Bechtel (*1995: 125). More generally, see Gottschalk and Gleser (1969), Gottschalk (1979), Gottschalk et al. (1986).

57. Even a 100% valid automated content analysis would not make obsolete the use of rhetoric in content analysis.

58. I am grateful to my daughter Marianna for bringing Farnham's work to my attention.

59. On the uses of content analysis, see for all Kassarjian (1977: 10), Neuendorf (2002: 191–213).

60. Another survey article of articles on advertising by Yale and Gilly (1988), also based on content analysis of ten years of six major advertising/marketing journals from 1976 to 1985 (for a total of 907 journal articles), shows that content analysis is used by an average of 7% of the articles across all journals (but with a peak of 18% in the *Journal of Marketing*). Unfortunately, the analyses carried out by Yale and Gilly provide less useful information on the way content analysis is used in these articles.

61. This is a consistent finding in empirical studies of content analysis research. Reporting of intercoder reliability measures is not standard (e.g., Lombard et al., 2002).

62. Weber (*1983: 144).

63. See also Schwalbe (*2006: 270), King and Lester (*2005: 627). Sometimes, even hypotheses read more like descriptions (e.g., *Law and Peixoto Labre, 2002: 700, *Tak et al., 1997: 420).

64. See King's magisterial historical work (1995) on the role of the US federal government in maintaining racial separation and inequality.

65. See also Jackson (2007) for a similar use of content analysis to test the hypothesis that occupational positions are allocated on the basis of meritocratic criteria. La Noce et al.'s work (2006) on the consistency analysis of the decisions taken by the Italian Competition Authority on mergers and acquisitions (and where content analysis is based on Bolasco's semi-automatic approach, 2005) is similarly interested in testing a decision-making model. Hobolt and Klemmemsen's (2005) estimate a formal model of governments' policy responsiveness. In these studies, content analysis provides only some but not all of the data.

66. For the peculiar problems of doing content analysis on the internet, see McMillan (* 2000).

67. The similarities are made even more apparent when content analysis categories are worded as questions (e.g., *Semetko and Valkenburg, 2000).

68. Untimely conclusion, in retrospect, at least. There is no reason to believe that Janowitz should be aware of the publication of Holsti's or Gerbner et al.'s books in the same year of the publication of his article.

69. See Griswold (*1981: 763), Law and Peixoto Labre (*2002: 702), Bauer et al. (*2006: 106), Katz et al. (*1969: 451).

70. On this issue, see Hruschka et al. (*2004: 321–322).

Bibliographical References

Abell, Peter. 1987. *The Syntax of Social Life: The Theory and Method of Comparative Narratives.* Oxford: Clarendon Press.

Atkinson, Paul (ed.). 2006. *Narrative Methods.* Benchmarks in Social Research Methods. Beverly Hills, CA: Sage.

Berelson, Bernard. 1952. *Content Analysis in Communication Research.* Glencoe, IL: The Free Press.

Berelson, Bernard. 1979. "Douglas Waples, 1893–1978." *Library Quarterly*, Vol. 49, No. 1, pp. 1–2.

Bolasco, Sergio. 2005. "Statistica testuale e text mining: alcuni paradigmi applicativi." *Quaderni di Statistica*, Vol. 7, pp. 1–37.

Budd, Richard W., Robert K. Thorp, and Lewis Donohew. 1967. *Content Analysis of Communications.* New York: Macmillan.

Burke, Kenneth. 1957. *Philosophy of Literary Form: Studies in Symbolic Action.* New York: Random House.

Corsaro, William A. and David R. Heise. 1990. "Event Structure Models from Ethnographic Data." *Sociological Methodology*, Vol. 20, pp. 1–57.

de Sola Pool, Ithiel (ed.). 1959. *Trends in Content Analysis.* Urbana: University of Illinois Press.

Dijk, Teun A. van (ed). 2007. *Discourse Studies.* Benchmarks in Social Research Methods. Beverly Hills, CA: Sage.

Dovring, Karin. 1954. "Quantitative Semantics in 18th Century Sweden." *Public Opinion Quarterly*, Vol. 18, pp. 389–394.

Drew, Paul (ed.). 2006. *Conversation Analysis.* Benchmarks in Social Research Methods. Beverly Hills, CA: Sage.

Farnham, Christie Anne. 1994. *The Education of the Southern Belle: Higher Education and Student Socialization in the Antebellum South.* New York: New York University Press.

Franzosi, Roberto. 2004a. *From Words to Numbers: Narrative, Data, and Social Science.* Cambridge: Cambridge University Press.

Franzosi, Roberto. 2004b. "Content Analysis." In: pp. 547–66, Alan Bryman and Melissa Hardy (eds), *Handbook of Data Analysis.* Beverly Hills, CA: Sage.

Geller, A., D. Kaplan, and Harold D. Lasswell. 1942. "An Experimental Comparison of Four Ways of Coding Editorial Content." *Journalism Quarterly*, Vol. 19, No. 3, pp. 362–370.

Gerbner, George, Ole R. Holsti, Klaus Krippendorff, William J. Paisley, and Philip J. Stone (eds). 1969. *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques.* New York: John Wiley & Sons.

Gottdiener, Mark (ed.). 2003. *Semiotics.* Benchmarks in Social Research Methods. Beverly Hills, CA: Sage.

Gottschalk, Louis A. (ed). 1979. *The Content Analysis of Verbal Behavior: Further Studies.* New York: Spectrum Publications.

Gottschalk, Louis A., Fernando Lolas, and Linda L. Viney (eds). 1986. *Content Analysis of Verbal Behavior: Significance in Clinical Medicine and Psychiatry.* Berlin: Springer-Verlag.

Gottschalk, Louis A. and G.C. Gleser. 1969. *The Measurement of Psychological States through the Content Analysis of Verbal Behavior.* Berkeley, CA: University of California Press.

Green, Bert F., Jr. 1967. "The General Inquirer: A Computer Approach to Content Analysis. Book Review." *American Educational Research Journal*, Vol. 4, No. 4. pp. 397–398.

Grier, David Alan. 2005. *When Computers Were Human.* Princeton: Princeton University Press.

Hall, Calvin S. and Robert L. Van De Castle. 1966. *The Content Analysis of Dreams.* New York: Appleton-Century-Crofts.

Heise, David R. 1989. "Modeling Event Structures." *Journal of Mathematical Sociology*, Vol. 14, pp. 139–69.

Heise, David R. 1989. "Modeling Event Structures." *Journal of Mathematical Sociology*, Vol. 14, pp. 139–69.

Hobolt, Sara B. and Robert Klemmemsen. 2005. "Responsive Government? Public Opinion and Policy Preferences in Britain and Denmark," *Political Studies*, Vol. 53, No. 2, pp. 379–402.

Holsti, Ole. 1969. *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison Wesley.

Jackson, Michelle. 2007. "How Far Merit Selection? Social Stratification and the Labour Market." *British Journal of Sociology*, Vol. 58, No. 3.

Janis, Irving L. 1943. "Meaning and the Study of Symbolic Behavior." *Psychiatry: Journal of the Biology and the Pathology of Interpersonal Relations*, Vol. 6, pp. 425–439.

Janis, Irving L. 1949. "The Problem of Validating Content Analysis." In: pp. 55–82, Lasswell, Harold D., Nathan Leites and Associates. *Language of Politics: Studies in Quantitative Semantics*. Cambridge, Mass.: The MIT Press.

Janis, Irving L. and Raymond H. Fadner. 1942. "A Coefficient of Imbalance for Content Analysis." *Experimental Division for the Study of War Time Communications*, Document No. 31, Nov. 1. Washington, DC: Library of Congress.

Janowitz, Morris. 1969. "Harold D. Lasswell's Contribution to Content Analysis." *Public Opinion Quarterly*, Vol. 32, No. 4, pp. 646–653.

Kassarjian, Harold H. 1977. "Content Analysis in Consumer Research." *Journal of Consumer Research*, Vol. 4, No. 1, pp. 8–18.

King, Desmond. 1995. *Separate and Unequal: Black Americans and the US Federal Government*. Oxford: Oxford University Press.

Krippendorff, Klaus. 1980. *Content Analysis. An Introduction to its Methodology*. Beverly Hills, CA: Sage.

La Noce, Mauro, Sergio Bolasco, Elisabetta Allegra, Valerio Ruocco, and Federico M. Capo. 2006. "Merger Control in Italy 1995–2003: A Statistical Study of the Enforcement Practice by Mining the Text of Authority Resolutions." *International Journal of the Economics of Business*, Vol. 13, No. 2, pp. 307–334.

Lasswell, Harold D. 1938. "A Provisional Classification of Symbol Data." *Psychiatry: Journal of the Biology and the Pathology of Interpersonal Relations*. Vol. 1, pp. 197–204.

Lasswell, Harold D. and Associates. 1942. "The Politically Significant Content of the Press: Coding Procedures." *Journalism Quarterly*, Vol. 19, No. 1, pp. 12–23.

Lasswell, Harold D., Ralph D. Casey, and Bruce Lannes Smith. 1935. *Propaganda and Promotional Activities; An Annotated Bibliography*. Minneapolis: University of Minnesota Press.

Lasswell, Harold D., Daniel Lerner, and Ithiel de Sola Pool. 1952. *The Comparative Study of Symbols*. Stanford, CA: Stanford University Press.

Laver, Michael. 1998. "Party Policy in Britain, 1997: Results from an Expert Survey." *Political Studies*, Vol. 46, No. 3, pp. 336–347.

Lebart, Ludovic. 1993. "Sur les analyses statistiques de texts." *Journal de la societé statistique de Paris*, Vol. 134, No. 4, pp. 17–36.

Lebart, Ludovic and Andre Salem. 1994. *Statistique textuelle*. Paris: Dunod.

Lebart Ludovic, Andre Salem, and Lisette Berry. 1998. *Exploring Textual Data*. Dordrecht: Kluwer Academic Publishers.

Lieberson, Stanley. 1985. *Making it Count. The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability." *Human Communication Research*, Vol. 28, No. 4, pp. 587–604.

Namenwirth, Zvi J. 1973. "Wheels of Time and the Interdependence of Value Change in America." *Journal of Interdisciplinary History*, Vol. 3, No. 4, pp. 649–683.

Namenwirth, Zvi J. and Robert Philip Weber. 1987. *Dynamics of Culture.* Boston: Allen & Unwin.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook.* Beverly Hills, CA: Sage.

Psathas, George. 1967. "The General Inquirer: Useful or Not?" *Language Resources and Evaluation*, Vol. 2, No. 1, pp. 12–16.

Richardson, John V., Jr. 1980. "Douglas Waples (1893–1978): A Biographical Sketch." *Journal of Library History*, Vol. 15, No. 1, pp. 76–83.

Roberts, Carl W. 1997a. "A Generic Semantic Grammar for Quantitative Text Analysis: Applications to East and West Berlin Radio News Content from 1979." *Sociological Methodology*, Vol. 27, pp. 89–129.

Roberts, Carl W. 1997b. "A Theoretical Map for Selecting Among Text Analysis Methods." In: pp. 275–285 Carl W. Roberts (ed), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Rogow, Arnold A. (ed). 1969. *Politics, Personality, and Social Science in the Twentieth Century: Essays in Honor of Harold D. Lasswell.* Chicago: University of Chicago Press.

Rosengren, Karl Erik (ed). 1981. *Advances in Content Analysis.* Beverly Hills, CA.: Sage.

Rosten, Leo. 1969. "Harold Lasswell: A Memoir." In: pp. 1–14, Arnold A. Rogow (ed), *Politics, Personality, and Social Science in the Twentieth Century: Essays in Honor of Harold D. Lasswell.* Chicago: University of Chicago Press.

Schrodt, Philip A. 2006. "Twenty Years of the Kansas Event Data System Project." *The Political Methodologist*, Vol. 14, no. 1, pp. 2–6.

Shapiro, Gilbert and John Markoff. 1997. "A Matter of Definition." In: pp. 9–31, Carl W. Roberts (ed), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Sills, David L. 1981. "Bernard Berelson: Behavioral Scientist." *Journal of the History of the Behavioral Sciences*, Vol. 17, No. 3, pp. 305–331.

Smith, Bruce Lannes, Harold D. Lasswell, and Ralph D. Casey. 1946. *Propaganda, Communication, and Public Opinion. A Comprehensive Reference Guide.* Princeton, NJ: Princeton University Press.

Sorokin, Pitirim A. 1933. "Recent Social Trends: A Criticism." *Journal of Political Economy*, Vol. 41, No. 2, pp. 194–210.

Starosta, William J. 1984. "Qualitative Content Analysis: A Burkean Perspective." In: pp. 185–194, Gudykunst, William B. and Young Yun Kim (eds). *Methods for Intercultural Communication Research.* Beverly Hills, CA: Sage.

Starosta, William J. 1988. "A National Holiday for Dr. King?: Qualitative Content Analysis of Arguments Carried in the Washington Post and New York Times." *Journal of Black Studies*, Vol. 18, No. 3, pp. 358–378.

Stone, Philip J. 1997. "Thematic Text Analysis: New Agendas for Analyzing Text Content." In: pp. 35–54, Carl W. Roberts (ed), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Stone, Philip J., Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge, Mass.: MIT Press.

Sylvan, David J. 1991. "The Qualitative–Quantitative Distinction in Political Science." *Poetics Today*, Vol. 12, No. 2, pp. 267–286.

Tukey, John. 1977. *Exploratory Data Analysis.* Reading, MA: Addison Wesley Publishing Co.

Viney, Linda L. 1986. "Assessment of Psychological States through Content Analysis of Verbal Communications." In: pp. 55–84, Gottschalk Louis A., Fernando Lolas, and Linda L. Viney (eds), *Content Analysis of Verbal Behavior: Significance in Clinical Medicine and Psychiatry.* Berlin: Springer-Verlag.

Waples, Douglas, Bernard Berelson, and Franklyn R. Bradshaw. 1940. *What Reading Does to People: A Summary of Evidence on the Social Effects of Reading and a Statement of Problems for Research.* Chicago: University of Chicago Press.

Waples, Douglas and Bernard Berelson. 1941. "What the Voters Were Told (An Essay in Content Analysis) (mimeographed)". Graduate Library School, University of Chicago.

Weber, Robert Philip. 1990. *Basic Content Analysis.* Beverly Hills, CA: Sage.

Whewell, William. 1837. *The History of the Inductive Sciences, from the Earliest to the Present Times.* London: J.W. Parker.

Whewell, William. 1840. *The Philosophy of the Inductive Sciences, Founded upon Their History.* London: J. W. Parker.

Willey, Malcolm Macdonald. 1926. *The Country Newspaper. A Study of Socialization and Newspaper Content.* Chapel Hill: University of North Carolina Press.

Woodward, Julian L. 1930. *Foreign News in American Morning Newspapers. A Study in Public Opinion.* New York: Columbia University Press.

Woodward, Julian L. 1934. "Quantitative Newspaper Analysis as a Technique of Opinion Research." *Social Forces*, Vol. XII, pp. 526–37.

Yale, Laura and Mary C. Gilly. 1988. "Trends in Advertising Research: A Look at the Content of Marketing-Oriented Journals from 1976 to 1985." *Journal of Advertising*, Vol. 17, No. 1, pp. 12–22.