CHAPTER 5

CONTENT ANALYSIS

Content analysis is the longest established method of text analysis among the set of empirical methods of social investigation (Holsti 1968, Silbermann 1974, Herkner 1974). It is for the present, however, somewhat difficult to proceed on the basis of a homogeneous understanding of method, in view of the rich and varied literature on 'content analysis'. Originally the term referred only to those methods that concentrate on directly and clearly quantifiable aspects of text content, and as a rule on absolute and relative frequencies of words per text or surface unit. Subsequently the concept was extended to include all those procedures which operate with (syntactic, semantic or pragmatic) categories, but which seek at least to quantify these categories by means of a frequency survey of classifications.

The establishment of 'qualitative content analyses' (Mayring 1988) has made it difficult to separate these from other methods of text analysis, particularly those oriented towards ethnographic methods or grounded theory. It has become clear 'that the range of procedures in content analysis is enormous, in terms of both analytical goals and the means or processes developed to pursue them' (Merten 1983: 46). If one were to accept this interpretation, one could describe as variants of content analysis all those methods of text analysis which somehow approach texts by means of categories, since it is no longer a matter only of the communicative content of texts but also of their (linguistic) form. In content analysis it is, therefore, more a question of a research strategy than of a single method of text analysis. We shall endeavour in what follows to present the fundamentals of classical, quantitative content analysis, and also to give a typology of content analysis procedures in the broader sense, with 'qualitative variants' as a further development.

5.1 THEORETICAL ORIGINS

The development of content analysis has, in essence, been influenced by the development of mass media and by international politics. If one discounts the early work in psychoanalysis (Freud's interpretation of dreams), then content analysis has grown in significance – particularly with the meteoric expansion of mass communication – in the first half of the twentieth century. Berelson

(1952), in his keynote work *Content Analysis in Communication Research*, argues that in the United States between 1921 and 1930 only 10 or 15 content analyses were carried out. The first noteworthy activity – brought about by the rapid development of the press, cinema and radio – was seen in the second half of the 1930s (Silbermann 1974: 254).

The theoretical basis of this first move towards analyses of content was Harold D. Lasswell's model of mass communication: Lasswell's formula 'who says what to whom and with what effect' determined the course of research in modern mass communication. Interest has consequently focussed communicator, receiver and communicative effect, where a clear causal interrelation is assumed. To investigate this interrelation the communicative content had to be quantified as precisely as possible. Lasswell succeeded in establishing the method by emphasizing above all the political value of content analysis. (For work on war propaganda see Smith et al. 1946.)

The development of content analysis was marked quite essentially by three conferences (see Merten 1983: 41ff.). In August 1941 the first conference on interdisciplinary mass communication research took place in Chicago, and – among others – Harold D. Lasswell, Bernhard Berelson and Paul Lazarsfeld took part. During this conference, not only was the term 'content analysis' coined for the new method but Lasswell, in a keynote address, set out the approach and the goals of content analysis: signs and statements are analysed to test their effect on the audience; the results are the frequencies of particular symbols, their intensity and the assessment of the senders. A second conference was organized by Charles E. Osgood at Allerton House (Monticello, Illinois) in 1955, during which qualitative and quantitative approaches were presented. New theories – such as Shannon and Weaver's (1949) 'information theory' – and refined analytical techniques (Osgood's 'contingency analysis' and 'evaluative assertion analysis' [Osgood et al. 1954, Osgood 1959]), together with Bales's (1950) 'Interaction Process Analysis' led to a situation where content analysis was understood as 'communication analysis' and its preoccupation with printed texts gradually diminished. In 1967 a third content analysis conference took place at Annenberg School in Philadelphia and here the scope of the method was discussed. From a theoretical viewpoint, previous analyses were criticized on the grounds that the results of content analysis must remain ambivalent as long as the investigator is not explicit about the selection procedure which is imposed by the context of some content and by the goals of the analysis (Krippendorff 1969: 70f.).

According to Merten (1983: 45) further discussion of content analysis has the following characteristics: the structure and selectivity of communication processes, the development of perceptible indicators and multivariate techniques of analysis, the improvement of systems of notation through inclusion of the non-verbal domain, and the improvement of data analysis through the development of dedicated electronic text analysis packages. 'The long overdue debate with linguistics, whose stock of experience of text classification and text analysis has hitherto been completely ignored by Content Analysis, is only slowly coming into existence' (Merten 1983: 45).

5.2 BASIC THEORETICAL ASSUMPTIONS

In the early days of content analysis research there was unquestionably a simple behaviourist-oriented stimulus-response model of communication which set up an asymmetrical relationship between sender, stimulus and recipient. Content was viewed as the result of a communication process structured in accordance with Lasswell's classic formula: 'Who says what in which channel to whom and with what effect' (Lasswell 1946: 37). 'It is no accident that we are concerned here with a model of mass communication, which views communication as the transportation of a container known as "content", which is transported by a communicator through some medium to a recipient' (Merten 1983: 56ff.).

Morris's (1938, 1946) semiotic works promoted the recognition that communication happens not on the basis of stimuli but on that of the meanings which are attributed to them and which cannot be accessed through the concrete form of a stimulus. Content analysis, on the other hand, was influenced more by the news transmission model of Shannon & Weaver (1949: 7), even though this was explicitly intended only for the syntactic level: an information source ('sender') conveys, via a transmitter, some news as a signal which goes to some receiver and is forwarded to its destination. During transmission there are a range of possible sources of interference. Schramm (1954) adapted this news transmission model as a general communication model and (mis)interpreted the syntactic decoding as semantic decoding of content (Merten 1983: 74). 'This simplification of the attribution of meaning and of the whole process of communication was simply taken over for the purposes of content analysis and may well have contributed significantly to the image of content analysis as an objectivisable mechanism for the analysis of signs' (Merten 1983: 74, with reference to Herkner 1974: 167, and Lisch & Kriz 1978: 32).

5.3 OBJECTIVES OF THE METHOD

The objectives which may be pursued by methods of content analysis may be understood with reference to the following list of quotations, arranged in chronological order:

> In content analysis we look upon statements and signs as raw materials to be summarized in order to bring out either (1) the impact of content upon audience, or (2) the influence of control upon content. (Lasswell 1941, quoted from Lasswell 1946: 90)

> Content Analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication. (Berelson 1952: 18)

> Content Analysis is any research technique for making inferences by systematically and objectively identifying specified characteristics of messages. (Holsti 1968: 601)

> The classification of symbolic material by scientifically trained observers who should judge, with the assistance of explicit classification and procedural rules, which parts of the textual material fall within the categories of the research schema, and are truly characteristic of the available content analyses. (Ritsert 1972: 17)

5.4 OUTLINE OF THE METHOD

5.4.1 The procedures, instruments and rules of classical Content Analysis

5.4.1.1 Sampling

It will be possible in only a very small number of cases to investigate all the material relevant to a particular problem. As an alternative to an ideal total treatment, samples may be used based on the probability method, and under certain circumstances quota samples (cf. Merten 1983: 280ff.). Holsti (1968: 653ff.) recommends a multi-stage process of selection: (a) selection of sender, (b) selection of documents, and (c) selection of a subset of the documents.

5.4.1.2 Units of analysis

The units of analysis are the smallest components of texts in which the occurrence and the characterization of variables (properties, categories) are examined. Since a text does not consist of 'natural units', these have to be defined at the syntactic or semantic level for every concrete investigation (Herkner 1974: 173): (a) syntactically defined units are, for example, sign (word), sentence, complete text, area and time; (b) semantically defined units are, for example, person, statement and unit of meaning.

Holsti (1968: 647f.) distinguishes between recording units and context units: (a) the recording unit is the smallest textual unit within which the occurrence of variables is examined; (b) the context unit is invoked to establish the characterization of variables, such as their positive or negative assessment.

5.4.1.3 Categories and coding

The core and the central tool of any content analysis is its system of categories: every unit of analysis must be coded, that is to say, allocated to one or more categories. Categories are understood as the more or less operational definitions of variables. Any definition of categories should be explicit, complete and adequate (Herkner 1974: 174). In the process of coding it is recommended that every category should also be illustrated with textual examples which are subsequently taken as given and which facilitate the allocation of further textual units. Programmes that relate units of analysis to codes (the so-called 'code-and-retrieve' programmes) can be helpful in this process (Weitzman & Miles 1995: 148ff.).

The system of categories – in contrast to ethnographically or GT-oriented analyses – should be established before coding is undertaken. If it becomes clear

during the coding process, however, that a modified system of categories would be preferable – because some categories are missing or ambiguous, or are simply never used – then the entire textual material must be re-coded using the new categories.

In most cases categories are conceived as nominal scales: a unit of analysis either belongs or does not belong to this category. In principle, however, higher levels of scales are possible. The categories of a variable must fulfil the usual requirements: they must be mutually exclusive and complete (Herkner 1974: 175).

The system of categories, therefore, endeavours to operationalize the variables of the particular research question and thereby focuses on the research question or the hypotheses derived from it. Consequently, it is possible to set up both inductive schemata of categories, using some previous text interpretation (for example ethnographic or GT-oriented, cf. Chapters 6 and 7), and deductive schemata deriving from established theories. The desire for integrated systems of categories has often been expressed (see Herkner 1974: 175), but methods which use such standardized tools are not suited to all research questions. For particular areas there are well-established systems of categories, such as:

- for the analysis of interactions, Bales's (1950) interaction process analysis (IPA) and the SYMLOG method (*S*ystem for *M*ultiple *L*evel *O*bservation of *G*roups) of Bales and Cohen (1979);
- for the analysis of attitudes, the evaluative assertion analysis of Osgood et al. (1954); and
- for the analysis of achievement motivation, the method of McClelland et al. (1953).

Holsti (1968: 645), following Berelson (1952: 147ff.), formulates a list of types of category which may be used as the basis for the design of a system of categories:

- Subject, theme: what is it about?
- Direction: how is the theme dealt with?
- Norms: what is the basis for classification and evaluation?
- Values: what attitudes, goals and wishes are displayed?
- Means: what means are used to achieve the goals?
- Features: what features are used in the description of persons?
- Actors: who initiates particular actions and who carries them out?

- Authority: under what name are statements made?
- Origin: where did the communication come from?
- Goal: to whom is it directed?
- Place: where do the actions take place?
- Conflicts: what is the cause of any conflict? Who are the participants? How strong is the conflict?
- Outcome: is the end of the conflict happy, tragic or uncertain?
- Time: when does the action take place?
- Form or communication type: what channel of communication is used?
- Form of statement: what grammatical and syntactic forms can be discovered?
- Methods: what rhetorical or propaganda methods are employed?

This list shows great similarities to ethnographic question lists and shows that even classical content analysis does not confine itself purely to explicit communicative contents. Unlike ethnographic procedures (see Chapter 7) these questions are not answered directly by reference to a text, but form the basis for the development of a tool, that is a schema of categories.

5.4.1.4 Coding and reliability

When the schema of categories has been stated precisely, the coding process begins. Here the units of analysis are identified and allocated to categories. In order to ensure that a coder is using the same criteria for allocation of units of analysis and categories throughout the operation, and is not modifying the definitions of categories (intra-coder reliability), it is advisable that regular operational discussions are held (Herkner 1974: 176). To achieve an acceptable level of intra- and inter-coder reliability (agreement between different coders), there also need to be explicit definitions of the categories based on numerous examples and coder-training sessions, using material related to the text to be analysed. Inter-coder reliability can be assessed using a number of different measurements which indicate the relative proportion of units of analysis allocated to the same category by two different coders (cf. Herkner 1974: 177f., Lisch & Kriz 1978: 88ff., Merten 1983: 302ff.).

5.4.1.5 Analysis and evaluation

Frequencies and indices The simplest type of evaluation consists of counting the number of occurrences per category: here some relationship is assumed between frequency of content and meaning. The unconditional acceptance of this assumption was one of the principal causes of the disagreement between

Berelson (1952) and Kracauer (1952) who took a 'qualitative' standpoint. It is also customary to use different indices which correlate two separate measurements (Herkner 1974: 179f.). For purely syntactic indices this may be exemplified by the type-token ratio (quotient of the total of different words and the total word count) and the action quotient (quotient of the number of verbs and the number of adjectives). An example of a semantic index is provided by the discomfort-relief-quotient (total number of words indicating an unpleasant condition out of the total number of words indicating either a pleasant or an unpleasant condition).

Contingencies In contingency analysis it is not only the frequency which is investigated but also mutual dependency of variables. It is a question of whether the probability of a particular type of phenomenon (for example two themes) is more than randomly high or low.

More complex procedures Here grammatical and semantic aspects are examined for explicit rules. An example of this is evaluative assertion analysis (Osgood et al. 1954), where the attitude of a sender to particular persons or facts is investigated. This method comprises a standardized, scalar schema of categories and precise rules for coding and evaluation (for more detailed discussion see Herkner 1974: 181f., Merten 1983: 192ff.).

Additional multivariate analyses are available, based on the results of content analysis, and these are dependent on the scalar level of the variables. Evaluation must take account of problems of inference, both from the selected material to the total material, and also from the selected material to the senders, receivers or the communicative situation (cf. Merten 1983: 107ff., Herkner 1974: 183ff.).

5.4.1.6 A typology of content analysis procedures

Merten (1983: 115ff.), using the criteria 'analytical goals' and 'tools of analysis', attempts to provide a typology of content analysis procedures – where communicators, recipients and situational orientation belong to analytical goals, and semiotic levels (confusingly) belong to methods of analysis (Merten 1983: 101ff.).

- At the syntactic level we find analysis of such features as letters, syllables, words or sentences and their structures, in so far as these are purely formal.
- At the syntactic-semantic level it is a question of the influence of syntactic structures on meaning-formation.
- At the semantic level the meanings of words, sentences and so on are analysed.
- At the syntactic-pragmatic level there is an attempt to justify a relationship between syntax and textual effect.
- The semantic-pragmatic analysis seeks to relate this effect to particular meaning-bearing words or sentences.

- Finally, the purely pragmatic analysis looks for the truly pragmatic structures which govern the reception of the text (for example rhetorical structures).

Herkner (1974: 165), following Holsti (1968), classifies content analysis procedures according to the following features: purpose of investigation, semiotic level, type of comparison and research question (using Lasswell's formula). If one tries to combine these two taxonomies, the content analysis procedures found in the relevant literature may be classified on the basis of semiotic levels and research questions (see Figure 5.1).

5.4.2 Qualitative content analysis

In the 1950s a controversy was already developing about research strategies in content analysis. Berelson (1952) was the first to put together the methods and goals of quantitative content analysis which had been developed up to that time, and these concentrated on assessment on the basis of frequency analyses. Kracauer (1952) reacted critically to this quantitative orientation because it neglected the particular quality of texts – their meaning content. Kracauer felt that particular attention had to be paid to the reconstruction of contexts. 'Patterns' or 'wholes' in texts could be demonstrated, not by counting and measuring their manifest contents, but by showing the different possibilities of interpretation of 'multiple connotations'. For Kracauer, categories are also of central importance: 'What counts alone in quantitative analysis is the selection and rational organization of such categories as condense substantive meanings of the given text, with a view to testing pertinent assumptions and hypotheses' (Kracauer 1952: 637f.). He preferred, however, to construct these categories with reference to latent contents and the reconstruction of context, and to take account of the meaning of particular instances. Nevertheless, Kracauer's suggestions represent rather a shift of emphasis than an independent method. (Also see Ritsert 1972: 14ff. for discussion of the controversy between qualitative and quantitative content analysis.)

More recently Mayring's (1988) qualitative content analysis has achieved popularity (see, for example, Lamnek 1989: 202ff., Mayring 1991) although its independence, compared to the classical model, has been questioned (Lamnek 1989: 213). Mayring has developed a sequential model and proposes, as far as goals are concerned, three distinct analytical procedures which may be carried out either independently or in combination, depending upon the particular research question:

1 The *summary* attempts to reduce the material in such a way as to preserve the essential content and by abstraction to create a manageable corpus which still reflects the original material (Mayring 1988: 53). For this the text is (a) paraphrased, (b) generalized or abstracted, and (c) reduced.
2 *Explication* involves explaining, clarifying and annotating the material (Mayring 1988: 68). As a first step (a) a lexico-grammatical definition is

Semiotic levels	Object	Examples of procedure	Research question						
			Who?	What?	How?	To whom?	Why?	What situation?	What effect?
Syntactic	Syntactic characteristics of message	• Author analysis (style analysis)	○		○				
		• Personality structure analysis	○		○		○		
Syntactic-semantic	Syntax and meaning-creation	• Word class analysis	○		○				
		• Syntactic complexity analysis	○		○			○	
Semantic	Meanings of words, sentences	• Theme analysis	○	○				○	
		• Contingency analysis	○	○					
		• Field of meaning analysis	○	○			○		
Syntactic-pragmatic	Syntax and effect of message	• Frequency readability analysis	○		○	○			○
		• Structural readability analysis	○		○	○			○
		• Impact analysis	○		○	○			○
Semantic-pragmatic	Meaning and effect of message	• Value analysis	○	○			○		
		• Attitude analysis (EAA)	○	○			○		
		• Motif analysis	○	○			○		
		• Personality structure analysis	○	○			○		
		• Intelligibility analysis	○	○		○			○
		• Objectivity analysis	○	○			○		
		• Semantic differential	○	○		○			○
		• Symbol analysis	○	○			○	○	
		• Reality analysis	○	○				○	○
		• Interaction process analysis	○	○		○	○	○	○
		• Attribution analysis	○	○	○		○	○	
Pragmatic	Effect of message	• Resonance analysis		○	○	○			○
		• Interview analysis	○	○		○	○	○	○

FIGURE 5.1 A typology of content analysis procedures

attempted, then (b) the material for explication is determined, and this is followed by (c) a narrow context analysis, and (d) a broad context analysis. The narrow context analysis incorporates the text (cotext) and this corresponds to the meaning of context used in conversation analysis (see Chapter 8), while the broad analysis includes additional information about the senders and the situation. (See also the distinctions between broad and local context, or macro- and micro-context in Chapter 2, 2.5.) Finally, (e) an explicatory paraphrase is made of the particular portion of text and (f) the explication is examined with reference to the total context.

3 *Structuring* corresponds more or less to the procedures used in classical context analysis and is also viewed by Mayring (1988: 75) as the most crucial technique of content analysis, the goal of which is 'to filter out a particular structure from the material'. Here the text can be structured according to content, form and scaling. The first stage (a) is the determination of the units of analysis, after which (b) the dimensions of the structuring are established on some theoretical basis and (c) the features of the system of categories are fixed. Subsequently (d) definitions are formulated and key examples, with rules for coding in separate categories, are agreed. (e) In the course of a first appraisal of the material the data locations are marked, and (f) in a second scrutiny these are processed and extracted. If necessary the system of categories is re-examined and revised, which necessitates a reappraisal of the material. (h) As a final stage the results are processed. (For further treatment of the process see Mayring 1988: 68.)

The central part of the process – structuring – is clearly derived from classical content analysis. Here, too, units of coding and evaluation are set up and arranged in a schema of categories.

The process of content analysis therefore consists of nine stages (Mayring 1988: 42ff.):

- determination of the material;
- analysis of the situation in which the text originated;
- the formal characterization of the material;
- determination of the direction of the analysis;
- theoretically informed differentiation of questions to be answered;
- selection of the analytical techniques (summary, explication, structuring);
- definition of the unit of analysis;
- analysis of the material (summary, explication, structuring);
- interpretation.

5.5 QUALITY CRITERIA

For 'classical' content analysis, Berelson's (1952) criteria of objectivity, systematicity and quantification are appropriate. The research strategy that is regularly pursued here is governed by the traditional criteria of validity and reliability, where the latter is a precondition for the former (and not vice versa). In particular, two specific problems of content analysis are discussed here: problems of inference and problems of reliability.

Problems of inference relate to the possibility of drawing conclusions, on the one hand, about the whole text on the basis of the text sample and, on the other hand, about the underlying (theoretical) constructs such as motives, attitudes, norms, etc., on the basis of the text. As a result, inference in content analysis confines itself only to specific features of external and internal validity: if the operationalization is valid, is there an (internal) fit between the constructs and indicators? If the process of measurement is successful, then is the sample (externally) representative of the totality?

In considering *problems of reliability* particular attention is paid to the trustworthiness of the coding. To what extent do different coders agree in the coding of the same text (inter-coder reliability)? How stable is the coding of the same coders (intra-coder reliability)? Particularly for the assessment of inter-coder reliability, a range of measures and indices has been developed which all attempt to express the number of identical codings in relation to the overall total of codings (for further discussion see Herkner 1974: 177f., Lisch & Kriz 1978: 88ff., Merten 1983: 302ff.). Herkner (1974: 178) recommends, for example, the reliability index π, which is calculated as follows:

$$(1) \quad \pi = \frac{P - P_e}{1 - P_e}$$

where

$$(2) \quad P_e = \sum_{l=1}^{k} P_{il} P_{jl}$$

P_e here represents the measure of agreement which could be expected on the basis of chance where P is the empirically established agreement and 1 is the maximum value of P. π is thus calculated as the quotient of the superiority of the actual agreement between coders over the chance value and the maximum possible superiority. The measure of purely chance agreement P_e is calculated in (2) as follows: the relative share P_{il} of coder i of the judgements in category 1 is multiplied by the relative share P_{jl} of coder j of the judgements in category 1. These probabilities are finally added up for all possible judgements k.

This index takes account of the empirical allocation of judgements to categories in a suitable form to measure chance agreement. It thus becomes possible not only to compare the empirically found and maximum possible agreements, but also to remove from both precisely this possibility of 'chance agreement'.

Krippendorff (1980: 158) formulates the following specific quality criteria for content analysis:

1 Validity:
   (a) material-oriented – semantic validity, sample validity;
   (b) result-oriented – correlative validity, prognostic validity;
   (c) process-oriented – construct validity.
2 Reliability:
   (a) stability;
   (b) replicability;
   (c) precision.

Semantic validity relates to the meaning reconstruction of the material, and is expressed in the appropriateness of the category definitions, the key examples and the rules for coders. Sample validity refers to the usual criteria for precise sampling. Correlative validity refers to the correlation with some external criterion (for example the results of other methods). Construct validity relates, for instance, to previous success with similar constructs, established models and theories, and representative interpretations. Stability refers to whether the same results are obtained in a renewed application of the analytical tool to the same text. Replicability is the extent to which the analysis achieves the same results under different circumstances, for instance with different coders. Finally, precision assumes stability and replicability and denotes the extent to which the analysis meets a particular functional standard (Mayring 1988: 96ff.).

Krippendorff (1980) suggests four possible sources of error that may lead to a lack of reliability: (a) Features of the units of evaluation – do the problem locations, where there is some disagreement about coding, differ systematically from other material? (b) Properties of individual categories – are instances of disagreement particularly common with particular categories? Do these categories have unclear definitions? (c) Differentiation of categories – are the distinctions between categories too fine? (d) Properties of the coders – if the lack of reliability cannot be attributed to (a) to (c), then the problem is usually with the coders and may perhaps be solved by more careful selection, more thorough training, shorter operating periods, etc.

5.6 PRECONDITIONS AND AREAS OF APPLICATION

Content analysis will always be used if communicative content is of greatest importance, if operational schemata of categories can be formulated in advance or if the analysis is concerned only with the lexicon of a text. If classical procedures are to be applied, there has to be a quite precisely formulated research question and, ideally, a set of interesting variables that can be encapsulated in the form of a hypothesis. There is an exception for those procedures which provide standardized schemata of categories and, thereby, also research goals.

Apart from the variants which confine themselves to simple word counts, content analyses are based on schemata of categories which must be predetermined in both quantitative and qualitative varieties.

The transcription requirements then depend on the variables under investigation. There is a tendency, however, for content analysis not to investigate non-verbal or para-verbal phenomena. Since inter-coder reliability is an essential criterion of quality for the results of text analysis, it is recommended that in the categorization process at least two independent coders should be involved in the coding of texts.

Whether and to what extent contextual information is required will depend upon the research strategy: in Berelson's classical (1952) methodology only the manifest textual content is of interest. The questions formulated by Holsti (1968: 645) as a basis for schemata of categories are indeed related to context but have to be answered only from textual content. It is difficult, however, to circumvent the influence of contextual knowledge on the researchers. Cotext plays a role in the sense that textual examples provide essential assistance in the coding process. In Mayring's (1988) explicatory procedure the analysis of both cotext and context is explicitly included.

For the processing of large quantities of text, computer programs are available. Depending on the unit of analysis a range of programs may be used: word counts and calculation of indices (for example Textpack), text retrievers (for example Wordcruncher), textbase managers (for example MAX), code and retrieve programs (for example WinMax, AQUAD) (see Weitzman & Miles 1995).

The catalogue of the different academic disciplines which use text analysis procedures covers the whole area of the social sciences and important areas of the humanities. As early as 1974 (163ff.), Herkner gave examples of applications in psychology, psychiatry, social psychology, sociology, communication studies, ethnology and literary studies. Examples of linguistic content analyses are to be found in Wodak (1981, 1984), and Wodak and Schulz (1986).

5.7 SIMILARITIES AND DIFFERENCES IN COMPARISON WITH OTHER METHODS

Ethnographic and grounded theory methods also work with categories that function as an analytical framework. Unlike these, however, the categorization processes of content analysis require that the categories be set up and operationalized in advance. Changes in the schema of categories during the coding process should only be made in exceptional circumstances. Ethnographic methods – and in particular grounded theory – postulate, in contrast, an inductive development of categories (concepts and indicators) on the basis of textual data. Moreover, these procedures often dispense with quantification and so the significance of individual categories is never operationalized by means of frequency of codings within these categories.

In addition, SYMLOG and also narrative semiotics may be classified as semantic-pragmatic content analyses. These procedures provide, in advance, specific research questions and schemata of categories derived from them. And if we go back to a broad definition of content analysis, even the procedures of critical discourse analysis may be seen as multidimensional and multi-stage content analyses. In any case, the techniques of content analysis may well be used within the framework of critical discourse analysis.

The ethnomethodological methods (MCD, conversation analysis) are clearly distinguished from content analysis since they dispense completely with data categorization. Differences between content analysis and functional pragmatics, and between objective hermeneutics and DTA all have a similar basis.

5.8 LITERATURE

The choice of primary literature is particularly difficult in such a well-established method as content analysis, which has been exhaustively discussed for several decades. Here we shall attempt to present a number of 'milestones' in the development of the method:

Bales, Robert F. (1950), *Interaction Process Analysis*, Cambridge: Addison-Wesley.

Bales develops here a clear example of semantic-pragmatic content analysis with broad objectives. The IPA laid the foundation of Bales's reputation as one of the leading exponents of group sociology and saw in SYMLOG (see Bales & Cohen 1979, and Chapter 10) a further development.

Berelson, Bernhard (1952), *Content Analysis in Communication Research*, New York: Hafner.

Berelson (1952) is the first comprehensive and synoptic work exclusively devoted to content analysis, and constitutes an important landmark in the development of the method.

Holsti, Ole R. (1969), *Content Analysis for the Social Sciences and Humanities*, Reading, MS: Addison-Wesley.

What Berelson (1952) achieved can also be said of Holsti; following the Annenberg Conference, the state of the art in content analysis at the beginning of the 1970s is summarized here.

Kracauer, Siegfried (1952), 'The Challenge of Qualitative Content Analysis', *Public Opinion Quarterly*, 16: 631–42.

The Kracauer-Berelson controversy is a landmark in the historical develop-

ment of content analysis. The trigger was this short article by Siegfried Kracauer, in which he presents some important objections to a purely counting and measuring orientation in content analysis.

Krippendorff, Klaus (1969), 'Models of Messages: Three Prototypes', in George Gerbner, Ole Holsti, Klaus Krippendorff, William J. Paisley & Philip J. Stone (eds), *The Analysis of Communication Content. Development in Scientific Theories and Computer Techniques*, New York: Wiley, 69–106.

In his contribution to this collection, which summarizes the findings of the Annenberg Conference, Krippendorff asks fundamental questions about the communication model of content analysis and exposes the selectivity of the treatment of information.

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

Krippendorff (1980) provides one of the first summative accounts of methods, in which his discussion of the quality criteria of content analysis is particularly significant.

Lasswell, Harold D. (1941), *Describing the Contents of Communication. Experimental Division for the Study of Wartime Communication*, Doc. No. 9, Washington, DC: Library of Congress.

Lasswell, Harold D. (1946), 'Describing the Contents of Communication', in Bruce L. Smith, Harold D. Lasswell and Ralph D. Casey (eds), *Propaganda, Communication and Public Opinion*, Princeton, NJ: Princeton University Press, 74–94.

This article by Harold D. Lasswell (first published 1941, reprinted with minor amendments 1946) offers good insight into the aims of the pioneers of content analysis.

Lazarsfeld, Paul, Berelson, Bernhard & Gaudet, Hazel (1955), *The People's Choice. How the Voter Makes up his Mind in a Presidential Campaign*, (2nd edn), New York: Columbia University Press.

This key investigation of electoral behaviour, which also formulates the 'two-step-flow' model of communication, includes one of the first exemplary applications of content analysis.

McClelland, David C., Atkinson, John W., Clark, Russell A. & Lowell, Edgar L. (1953), *The Achievement Motive*, New York: Appleton Century Crofts.

McClelland and his collaborators develop a theory of achievement motivation and design a schema of categories for content analysis to be used in motivation research.

Osgood, Charles E. (1959), 'The Representational Model and Relevant Research Methods', in Ithiel de Sola Pool (ed), *Trends in Content Analysis*, Urbana, IL: University of Illinois Press, 33–88.

Osgood, Charles E., Saporta, Sol & Nunnally, Jum (1954), *Evaluation Assertive Analysis*, Chicago, IL: University of Chicago Press.

In the idea of 'evaluation assertive analysis' Osgood and his team, following the Allerton House Conference, present a schema of categories for content analysis which proved to be very fruitful for a variety of research questions.

Schramm, Wilbur (1954), *The Process and Effects of Mass Communication*, Urbana, IL: University of Illinois Press.

Schramm's work presents the widely accepted communication-theory basis for classical content analysis. Schramm reformulates here Shannon & Weaver's (1949) technological information model for the sphere of mass communication.

Shannon, Claude E. & Weaver, Warren (1949), *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.

5.9 SECONDARY LITERATURE

5.9.1 Manuals

Content analysis is included in almost all method manuals, although admittedly it is losing its unique position in more recent editions. For instance, in the latest edition of the *Handbook of Social Psychology* (Lindzey & Aronson 1985) there is no entry for content analysis. In place of this the authors have included a more general account by Clark (1985), 'Language Use and Language Users', which also outlines linguistic theories and methods (such as speech act analysis, conversation analysis). It can even be seen from the heyday of secondary publications on content analysis – the 1970s and early 1980s – that the lifecycle of the method has passed its peak and that a phase of differentiation has now begun.

Herkner, Werner (1974), 'Inhaltsanalyse', in Jürgen von Koolwijk & Maria Wieken-Mayser (eds), *Techniken der empirischen Sozialforschung*, vol. 3, München: Oldenbourg, 158–91.

Herkner's contribution to this manual gives a comprehensive overview of the emphases, the theoretical foundations and the procedures employed in content analysis research.

Holsti, Ole R. (1968), 'Content Analysis', in Gardner Lindzey & Elliot Aronson (eds), *The Handbook of Social Psychology*, (2nd edn), vol. 2, *Research Methods*, Reading: Addison-Wesley, 596–692.

Holsti provides, almost as a summary of his methodological book (1969), an overview of essential questions and procedures used in content analysis.

Mayring, Philip (1991), 'Qualitative Inhaltsanalyse', in Uwe Flick, Ernst von Kardorff, Heiner Keupp, Lutz von Rosenstiel and Stefan Wolff (eds), *Handbuch Qualitative Sozialforschung*, München: Psychologie-Verlags-Union, 209–13.

Philip Mayring's contribution, which briefly outlines his 'qualitative content analysis', is included in this German manual of qualitative social research.

Silbermann, Alphons (1974), 'Systematische Inhaltsanalyse', in René König (ed.), *Handbuch der empirischen Sozialforschung*, vol. 4, Komplexe Forschungsansätze, Stuttgart: Enke, 253–339.

Alphons Silbermann's contribution is particularly impressive for its thorough presentation of the historical development of the method. As fields of application, Silbermann describes the analysis of cultural and social thought systems, literary analysis, the analysis of stereotypes and symbolic representations, and the use of content analysis in warfare and politics. An overview of content analysis research in mass communication is followed by a section on the further development of the method and a short account of techniques.

5.9.2 Other descriptions of methods

Lamnek, Siegfried (1989), *Qualitative Sozialforschung*, vol. 2, Methoden und Techniken, München: Psychologie-Verlags-Union, 202–13.

Lamnek gives a brief account of Marying's (1988) qualitative content analysis and contrasts this, as a genuinely qualitative method, with objective hermeneutics.

Lisch, Ralf & Kriz, Jürgen (1978), *Grundlagen und Modelle der Inhaltsanalyse*, Reinbek: Rowohlt.

Ralf Lisch and Jürgen Kriz attempt a state of the art description and a critique in which they concentrate on the methodological foundations and the problems of sampling and categorizing, together with questions of validity and reliability. Further sections deal with content analysis and computing, with measurement of direction and intensity, and with the statistical analysis of trends, association structures and readability research.

Mayring, Philip (1988), *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, Weinheim: Deutscher Studienverlag.

In this work Mayring gives the clearest presentation of the general conception of his qualitative content analysis. At the centre are the 'techniques' (which he calls the 'objectives') of qualitative content analysis: summary, explication and structuring.

Merten, Klaus (1983), *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*, Opladen: Westdeutscher Verlag.

Klaus Merten succeeds in this book in providing probably the fullest and most detailed German account of content analysis. In addition to matters dealt with in other method descriptions (history and development, theoretical principles, problems of inference), this book is characterized by a typology of very different procedures and by an attempt to update the communication theory basis and to integrate some more recent social science approaches (Niklas Luhmann).

Ritsert, Jürgen (1972), *Inhaltsanalyse und Ideologiekritik. Ein Versuch über kritische Sozialforschung*, Frankfurt: Athenäum.

Jürgen Ritsert is concerned with the question of the place value of content analysis in an ideologically critical research programme based on the critical theory of the Frankfurt school.

Weber, Robert Philip (1990), *Basic Content Analysis,* (2nd edn), Newbury Park, CA: Sage (Quantitative Applications in the Social Sciences Series).

Weber treats questions from the creation of a simple coding scheme to an elaborate computer-aided analysis of content. He makes his points with numerous well-chosen pieces of text: US political party platforms, and Korean War editorials from American newspapers. At the end of each chapter, he provides a useful discussion of current literature. He finishes with a sensitive discussion of the unresolved problems in measurement, indication, representation and interpretation.

Wersig, Gernot (1968), *Inhaltsanalyse. Einführung in ihre Systematik und Literatur. Schriftenreihe zur Publizistikwissenschaft*, Vol. 5, Berlin: Volker Spiess.

Gernot Wersig's publication is noteworthy for a comprehensive and systematic bibliographic overview of investigations in content analysis.

5.9.3 Examples of studies

In 1968 Gernot Wersig was already able to list 1,400 publications on content analysis. If one discounts the communication theory or methodologically focussed contributions, there still remain more than a thousand examples of applications of content analysis. At the very least we can claim that it is the method of mass communication research. It is therefore all the more difficult to highlight individual studies as particularly exemplary. Moreover, this task is made even more difficult by the complexity of content analysis procedures. The only remaining possibility is to point again to a number of classical studies that have already been discussed (for example Lazarsfeld et al. 1955, Osgood et al. 1954).

Examples of didactically–oriented applications of 'classical' content analysis may be found in Merten (1983: 312–28), while examples for qualitative content analysis are given in Mayring (1988).