PROBLEMS OF SCIENTIFIC DEFINITIONS OF DOCUMENT (CONTENT) ANALYSIS IN SOCIAL SCIENCES

Dragan Tančić[1], Petar Tančić[2], Lidija Tančić[3]
[1]Institut of Serbian Culture, Pristina/Leposavic, Serbia
[2]Faculty of Business Studies and Law, master student
[3]Faculty of Philology, Belgrade, student
dragan_tancic@yahoo.com

Abstract: *The method of document (content) analysis in social sciences is an important factor of a science, scientific knowledge. Scientific knowledge about this scientific method in the methodology of social and political sciences indicates that this is a highly complex and thorough scientific method, with which particular knowledge on social, political and other phenomena and processes is obtained not only from near, but also from distant past. In the existing scientific corpus there is no universal agreement as to the scientific definitions of this scientific method. In logic and methodology, document (content) analysis was unjustifiably considered to be just an assisting tool in the research of particular phenomena and processes..*

Keywords: *scientific method, methodology, document (content) analysis method, social and political sciences.*

1. INTRODUCTION

Understanding of the scientific method in social and political sciences is one of the relevant elements of the basic concept of a certain science, which consequently bears, along with research-operational, also constitutional importance for a science. In this context we can emphasize that it is not rare in a science that there are more scientific methods and various interpretations of the most suitable manner of organizing scientific field, which are derived from different general theoretical and meth-

odological aspects, different nature of a problem. The scientific method rules out all approaches that do not possess a scientific merit and are not scientifically grounded.

According to Carl Popper, a scientific method is a method of audacious and ingenious assumptions and sharp refutations thereof. An audacious assumption is a theory with a large content – larger than that of current theories, which will, hopefully, be left behind by it. Popper points out that a scientific method is not a route to success in a science and that it is not a way of justifying scientific results, but that scientific results can only be criticized and examined. (Popper, 1974)

Nagel believes that 'the application of a scientific method is a constant criticism of arguments in the light of tested criteria for evaluating the reliability of methods used for obtaining data, and in the light of criteria for the appraisal of the proving potential of the evidence which the conclusions are based upon. (Nagel, 1974)

We can generally point out that a scientific method has three fundamental elements:

1. Logical part, consisting of fundamental rules of logical an.

2. Epistemological part, a method is an essential structural part of every theory. All theories, scientific knowledge, have their research subjects which include natural, psychological and social phenomena. Obviously, even here we come across substantial differences most directly expressed within the categorical system. Each theory that refers to some of the mentioned types of phenomena has an established special conceptual system and its own symbolism and terminology. The same terms in different theories have different conceptual meaning. The analysis can show us that each of the mentioned types of phenomena is being further differentiated within itself, for each of them a special theory and a special categorical system is formed, a special conceptual and terminological apparatus. As opposed to the first part which is identical for all scientific fields, in this part differences are significant.

3. Methodological – technical part, contains methods and techniques (instruments and procedures) through which scientific knowledge is acquired.

In the methodology of political sciences there are various classifications of scientific methods, from which we will single out two, taking into account their scopes and significance in the methodology of social sciences.

The classification by the famous French author Madeleine Grawitz (1973) gives special significance to the historical method as one of the theoretical-methodological approaches in the research that is directed to the scientific explanation of social phenomena, which can be seen from the arrangement of these methods:

1) Comparative method,

2) Historical method,

3) Genetic method,

4) Functional method,

5) Structural method,

6) System analysis,

7) Dialectical method.

Klaus von Beyme (1974) also emphasizes the significance of a historical-genetic method, which can be noticed in the arrangement of 'methodological postulates of political scientific research':

1) historical-genetic,

2) institutional,

3) behavioristic,

4) functional-structural,

5) comparative.

Beyme points out that politicologists are necessarily directed to historical research, to the application of historical-genetic method, which has a special function in the field of science of history and science of politics, especially from the aspect of meaning of chronology, methods of quantitative biographical research and from the aspect of ideological-critical research.

In the methodology of social and political sciences the most often present classification criteria are based upon generality and method justification, applicability, relatedness of a research subject and methods to particular sciences, scientific disciplines, upon criterion of data collection, organization and processing, criterion of deduction on collected data, criteria of origins-derivation of theories, and upon other criteria.

General scientific methods are applied in the research of all types of phenomena, they are unequally applicable to different subject matters.

Basic methods (analysis-synthesis, induction-deduction, specialization-generalization, abstraction-concretization) are in a specific way differently applied in the research of phenomena of all sorts.

2. PROBLEMS OF DEFINING METHOD OF DOCUMENT (CONTENT) ANALYSIS

When considering the content analysis method in social and political sciences, we think that it is necessary to emphasize several relevant provisions:

1) the content analysis method as a scientific method is based upon scientific knowledge and it has been developed along with the development of scientific knowledge, and

2) the content analysis method as a method of scientific research, i.e. as a method (whose research project, research procedure, data processing and deduction are according to scientific norms and features) with which during research scientific knowledge on particular social and political phenomena and processes is truly acquired. In this context it is necessary to emphasize the following provisions as well:

a) the application of content analysis method has begun well before it was constituted as a method of scientific research and in a wider scope than that of scientific research;

b) content analysis method spontaneously applied, more or less successfully, by lots of subjects who do not possess appropriate technical and especially, scientific knowledge about it as a scientific method and method of scientific research;

The term Content Analysis is usual in the methodology of social and political sciences. When defining the concept of document (content) analysis in methodological, political, sociological and psychological literature, as well as in the literature of other scientific disciplines, different terms are mentioned. The most common are 'document analysis', 'document observation', 'document research' and 'document content analysis. The use of lots of terms points to the fact that in scientific theory there is no universal agreement as to the name of this method, and that a lot of issues related to it still remain unresolved. According to Warples and Berelson, by applying systematic content analysis, one makes effort to clarify supporting content descriptions for the purpose of providing objective account of the nature and relevant strength of encouragement that have effect on readers or listeners. (Waples, Berelson, 1941) Berelson argues that content analysis is a technique for researching objective, systematic and quantitative descriptions of obvious content of communication. (Berelson, 1952) Similar to Berelson, Holsti defines content analysis as a technique for objective and systematic description of manifest content of a text, wherein 'manifest' stands for the description of obvious (open), and not latent content. (Holsti, 1969) According to the Zagreb-based politicologist, Pavle Novosel, content analysis is a highly formal-

ized procedure, whose main goal is to break the qualified material into a string of units based on a specific categorical system and then process the units quantitatively. (Bibic, Novosel, 1971) Rudi Stojak pointed out '…that it is certainly not just a sum of methodological instruments and procedures since it is increasingly striving to get involved in the research fields which get to the core of the message, especially by means of qualitative description. The common factor for the whole content analysis technique is to reach the structure of the text by counting frequencies, extending and other operations in order to accomplish defunction and deduction.' (Stojak, 1990) Instead of the term content analysis, Slavomir Milosavljević and Ivan Radosavljević use the term document analysis and they define the method as "a necessary, thorough and reliable operative method for collecting and treating data on past, present and future phenomena, phenomena either close or very distant in space and time.' 'Any factor, part of or whole political phenomenon, its quantitative or qualitative properties, etc. if it is in any way recorded" can be a direct subject matter of research where content analysis is applied. (Milosavljević, Radosavljević, 2000)

In methodology and politicology, we usually talk about 'document content analysis'. The reason for using that term in politicological research is probably the fact that the method is primarily applied to document content. However, this term is inadequate for a number of reasons. The focus of it is exclusively on the content of documents, whereas its other properties are neglected: form, shape, structure, age, origin, etc. These properties are relevant providers of data and information in many fields of research, especially in historical science. The term 'document observation' is even less adequate, since, even though it is closer to sensory perception of documents, it neglects the thinking processes and the fact that knowledge about the subject matter of research is gained by inspecting verbal statements contained in various written documents, and, yet, these statements directly belong to the research. For that reason, the knowledge contained in a document is not directly acquired through sensory perception (empirically), but through the statements of the author of the document, which is truly an indirect written investigation.

These issues related to determining the name of the method make it obvious that the names used so far are not entirely adequate, and that we should look for a new name which will better reflect its essence. Hence the name "document content analysis" which entails the analysis of both content and form of a document, and both observation and investigation, seems more appropriate and precise to us. Namely, analysis is the basic method for gaining knowledge and scientific knowledge, but knowledge acquisition never ends with it.

The research subjects of Document (Content) Analysis Method are general, specific and individual events, social realities, states, properties, activities etc. which are investigated through statements contained in various documents, and their main properties are: coherence and meaning, as well as integrity and statement specification by time, space, content and form. From the point of view of social and political sciences, it is not particularly important whether the subject of document (content) analysis is researched directly (as the event itself) or indirectly (the presentation of the event in documents). What is important is the fact that not only events are investigated but also natural and social creations which constitute social realities, and not events.

Some authors believe that messages can be the subject of the document content analysis, which is basically true, since most coherent statements contain a message. However, it is unclear whether all coherent statements contain a message in the basic sense of the word, which makes message, as an important part of complex statements, too narrow a criterion for the subject of research. Meaning can also be the subject of document (content) analysis. Namely, there is no coherent statement without any meaning, but meaning is not the only component of a statement. In cases where meaning is the subject of document (content) analysis, many questions arise: What meaning are we talking about? Is it the directly expressed meaning or the meaning of the context? Who does the meaning refer to, etc.?

In the existing scientific corpus, there is mention of three types of document (content) analysis: classical, quantitative and qualitative analysis. In some papers, depending on the analysis criterion which is applied, the following types are also mentioned: external, internal, valent, statistical, dynamic, contingent, frequency, non-frequency, univariate, multivariate document (content) analysis.

Depending on the performance level, classical document (content) analysis can be: internal and external. It is thought of as a kind of qualitative – non-frequency content analysis, which was first developed by historians and later accepted by researchers from other scientific disciplines. Classical document analysis dates back to the origins of science and is divided into general analysis, which includes historical and sociological analysis, and special analyses, which are: legal, economic, psychological, political, statistical and other analyses. Taking into account the application of instrument properties criterion, it can be said that there are two main types of the document (content) analysis method: qualitative or non-frequency analysis and quantitative or frequency analysis. (Milosavljević, Radosavljević, 2002)

The number of authors who advocate the qualitative approaches in document

(content) analysis is not small. Among the most important are: P. Lazarsfeld, G. Ritzer, A. George, K. Baier, G. Gerbner, T. Adorno, K. Krippendorff, Z. Tanic, S. Milosavljevic and others. One of the leading authors in the field of sociological and political research methodology – Lazarsfeld, gave a significant contribution to distinguishing the oppositions between qualitative and quantitative approach to document (content) analysis. His main idea is based on the fact that behind the objective content in a document, there usually lies latent context, which is not visible, and therefore, it is necessary to see manifest and latent content in messages as equally important. Still, Lazarsfeld thought that in the application of content analysis, as a method, we should focus more on investigating complex, latent structures which cannot be measured directly, but about which we can make inferences based on objective contents in a message. Another advocate of the qualitative approach in content analysis is A. George, who believes that in content analysis we should investigate the presence and absence of a characteristic or a syndrome of the content as proposals for deduction. He rightfully emphasizes 'that when researching by the content analysis method, formal (frequency) characteristics of research must not be dominant and that, in research goals it is necessary to set basic hypotheses about certain content and then see whether that hypothetical content exists or doesn't exist in an adequate measure. (Stojak, 1990) Although he did not specifically deal with the content analysis method, in his remarks concerning the social research methodology, K. Baier points out 'that there are certain contents which are not quantifiable.' According to Baier, quantification is, in a way, "counting or aggregation of separate objects in which there is no continuous dynamic measuring.'(Baier, 1965) In an attempt to reconcile the contradictions between quantitative and qualitative approach in content analysis, G. Gerbner emphasizes 'that the primary task of a content analyst in the field of mass communication lies in the effort to gather various information about a certain content, and then test the excerpts from the content in order to arrive at objective conclusions on social phenomena.'Gerbner, 1964) Such understanding of the document (content) analysis method is a step further from Berelson's classical postulates and it is increasingly gaining the characteristics of a complex method for collecting data which serves to delve deeper into the original, hidden and unreachable information. In emphasizing the advantages of the qualitative content analysis, T. Adorno points to the need to understand 'meaningful connections'. He critically notices that 'words and sentences do not stand on their own in some stand-alone categories, therefore any opposite quantification is unsatisfactory.' (Adorno, 1962) However, although he does not accept the primacy of quantification and statistical semantics, he maintains

that 'science is measurement', but also adds 'that the qualitative content analysis dictates the social content of texts.' (Adorno, 1962)

A distinct representative of a "reconciler" of the qualitative and quantitative approach in content analysis is K. Krippendorff. He defines content analysis as 'a research technique for drawing repeated and valid conclusions from data, based on their context … which is determined on the basis of knowledge from previous research on a phenomenon.'(Krippendorf, 1981) From recent works in the field of document (content) analysis which have a balanced scientific approach towards qualitative and quantitative analysis, the following stand out: Rudi Stojak's 'Content Analysis Method' from 1990, Slavimir Milosavljevic and Ivan Radosavljevic's 'Basic Methodologies of Political Sciences' (2002); Dragan Tancic, Miodrag Gordic and Dzevad Termiz's 'Methodological Basis for Research on Safety, Defense and Terrorism' (2016:787).

From the above mentioned overview of authors who advocate the qualitative approach in investigating social phenomena, it is clear that there is no consensus as to the qualitative definition of the document (content) analysis. Even in the standard manual for content analysis by Lasswell, Leites at al., the qualitative document (content) analysis is not treated properly; it predominantly deals with the quantitative analysis. (Lasswell, Leites et all. 1949) That is why this methodological procedure – which is applied in the document (content) analysis – remains theoretically unexplained and practically undeveloped. That is one more reason for this paper to give more attention to the qualitative document (content) analysis. One of the first steps in that direction is an attempt to define that type of document (content) analysis.

In political sciences, qualitative analysis relates to the content of political documents. It is used for examining the contextual elements of content which is observed as a separate entity. In some cases, qualitative analysis boils down to merely stating that certain important elements of the content are either present or absent. That analysis does not investigate the prevalence or frequency of occurrence of certain categories, but only the quality of the manifest content and the manner in which it is explained in a given context. It is suitable for application on small samples or for analyzing individual documents (for example, a decision to call an election, a decision on accession of the state to "Partnership for Peace" and the like).

A number of authors have dealt with the qualitative document (content) analysis, and those who have given the biggest contribution to the development of that kind of content analysis method are: B. Berelson, H. Lasswell, A. Kaplan and J. Goldsten. For Berelson 'content analysis is a research technique for objective description, systematization and quantification of the manifest contents in communication.' (Barelson,

Layarsfeld, 1948) Such definition sprouted research procedures which were principally directed at discovering frequencies of some part of a message, symbol or other manifestation, which influenced researchers to develop statistical models of analysis, up to the use of computers in analysis. (Stojak, 1990; Mitchell, 2020) Berelson is responsible for the introduction of the term 'categories', thus further developing the methodological side of that method. He insisted that, during the categorization of a large material, special attention be given to important characteristics of categories, such as: homogeneity, completeness, mutual exclusion, objectivity, adequacy etc. (Stojak, 1990) Harold Lasswell gave the greatest contribution to the research of political propaganda using the content analysis method, and also to the study of structures of Nazi Germany's methods of propaganda in the Second World War. His main idea was that the propaganda messages could be a reliable source for predicting real war moves and plans. Lasswell and his associates defined content analysis as 'quantitative semantics' for which the statistical method is used, for the sake of objectivity and precision. (Lasswell, Lerner, de Pool, 1953) Kaplan and Goldsten emphasize the importance of quantitative characteristics in content analysis and claim 'that content analysis strives to quantitatively classify the content in the language of a category system set for providing data according to the specification of hypotheses which refer to that content.' (Kaplan, Goldsten, 1943) Kaplan gave another definition of content analysis in which he said 'that this method is trying to put focus on systematic and quantitative characteristic of meaning of a certain data set.' (Kaplan, 1943)

The question of which type of document (content) analysis method to use, is an artificial question. The document (content) analysis method is divided into qualitative and quantitative analysis in this paper for methodological reasons. Essentially, they are two aspects (approaches) of document content analysis which complement each other and are combined in research. Applying qualitative document (content) analysis in politicological research, the researcher discovers the general and basic sense of the message or information contained in a political document. Applying quantitative analysis, the researcher finds certain elements which can be counted or measured in political documents, in order to provide a stronger and more believable proof about the general and basic sense of the message or information contained in a political document. The researcher cannot see that message straight away from the qualitative analysis, but can only sense it, thus, quantitative analysis can be considered a higher level of hypothesis verification for which qualitative analysis cannot find a good enough proof. Therefore, in research we should strive for the synthesis of the qualitative and quantitative aspect of political document (content) analysis.

Without the qualitative aspect, quantitative political document (content) analysis makes no sense. Namely, it is necessary to add a corresponding political sense to obtained quantitative, i.e. numeral indicators.

For every kind and type of document (content) analysis, it is necessary to create a corpus of terms and codes (or, if possible, a corpus of messages, information and statements), wherein the corpus of terms is an essential element of document (content) analysis, and the corpus of codes is a technical and supporting part of the method. Making the corpus of terms and codes is among the most complex and most responsible jobs in applying the document (content) analysis.

The corpus of terms and codes is, first of all, a system of strictly defined basic categories and codes. Categories are seen as important basic words which are used to express certain social results, and codes are expressed through numbers – digits (numeric symbols), used for marking categories. Furthermore, the corpus of terms and codes is an important condition for using document (content) analysis as a scientific method for data collection. It is directly derived from indicators, the choice of which is determined by category selection, degree of their development, the relationship between them and similarities. For that reason, it can be said with certainty that the theoretical and operational development of subjects and hypotheses, as well as careful determination of research indicators, is the frame and the basis of corpus of terms and codes, i.e. categorical system. Namely, the basic connections, contexts and meanings are already contained in the subject matter and the hypotheses of the research, and their manifestations (concrete and abstract) – in the research indicators.

Many factors affect making of the corpus, its kind, type and properties. These are, first of all, the degree of development of science and its theory, which involves the degree of stability of its axioms, postulates and theorems, and the development of systems of categories, concepts and terms. Secondly, there are kinds of documents.[1] The third determinant of the process of making a corpus of terms and codes are subject matter, goals, purpose, hypotheses and research indicators.

Because of all that, in the process of making a corpus of terms and codes, it is extremely important to: a) develop the categorical system in great detail and, in it, strictly define and describe each concept, especially all the categorical concepts; b) develop standard expressions with specific meaning, standard figures and structures of statement which are most often used with its meanings; c) acknowledge time and space, i.e. the corpus should contain periodization with clear categorization and loca-

---

1   The requests for analyzing a statue, a bowl, a photograph, a theater piece, a movie, a television or a radio show, a newspaper article, etc. are uneven.

tion whose determination would be supplemented by meaning, and d) make detailed guidelines for its application in order to secure the proper decoding of the encoded material.

We can conclude that the corpus is an instrument for converting verbal statements to data, numerical data, and codes are digits used for marking categories, for the sake of easier data recording and manipulation, and statistical procedures application in data treatment and processing.

Since the corpus of terms and codes and the record sheet are methodological procedures and instruments of the document (content) analysis method, it is necessary to also point out their functions in research planning. The corpus of terms and codes and the record sheet are means of investigation, that is, they are ' tools'for collecting data in the document (content) analysis method. They are the final concretization of the scientific theory and through them, the scientific theory meets practice, i.e. the empirical reality of the researched (analyzed) phenomenon. The basic function of the corpus of terms and codes and the record sheet is to enable systematic and wholesome collecting, ordering and processing of data about the researched (analyzed) phenomenon.

3. CONCLUSION

In social sciences we can identify a number of scientific methods and different understandings and theoretical-methodological approaches to organizing scientific activities. The scientific method in social sciences excludes all the approaches which do not have scientific merit and are not scientifically grounded.

Based on its place in the research project and realization, the document content analysis method is an autonomous, empirical-theoretical method which can play many roles in research. On the one hand, it can be used to research the existing scientific document corpus before the development of the research project, it can be one of the methods for data processing, then it can be an autonomous method for obtaining and processing data and forming the scientific knowledge, and finally – the only possible method of the scientific knowledge in research.

The integral part of the document content analysis method is the corpus of terms and codes, which is directly derived from indicators, the choice of which also affects the choice of categories, the degree of their development, the mutual relationship and similarities of the document (content) analysis, especially the theory of the corpus

of terms and codes. The functions of the corpus are manifold: First, the corpus helps to systemize and verify current knowledge of the subject of research. Second, based on the corpus, the efforts of researchers are directed towards the central questions in the formulated subject of research. Third, the corpus serves as a basis for making the record sheet for data collection. Fourth, the corpus is the basis for phenomenon description, i.e. for the treatment of data and knowledge contained in a document. (Tančić, Gordić, Termiz, 2016:447-511) Corpus and record sheet development is an activity of equal importance as is the development of other parts of the scientific idea (subject, hypotheses, indicators, etc.) That is exactly why these instruments of the document (content) analysis method are conditioned (progressively) in almost every part of the scientific idea.

4. REFERENCE

[1] Kaplan, A. Goldsten, J. M. (1943) Reliability of certain Categories for classyfing Newspaper Headlines, Library of Congres.

[2] Kaplan, A. (1943) *Content Analysis and Theory of Sign,* Philos. Sciences, 1943.

[3] Berelson, B. (1952) Content Analysis in Communications Research, The Free Press, Glencoe, 1952, No. 37.

[4] Berelson, B. (1960) Analiza sadrzaja, Institute of Social Sciences, Belgrade

[5] Berelson, P. Layarsfeld (1948) *The Analysis of Communications Content,* University of Chicago, New York.

[6] Waples, D. Berelson, B. (1941) An Essay in Content Analisis, University of Chicago, Graduate Library School, 1941, No.2.

[7] Holsti, D. R. (1969) Content Analysis for the social Sciences and humanites, Addison Wesley

[8] Tancic, D. Gordic, M. Termiz, Dž. (2016) Oosnovi istrazivanja bezbednosti, odbrane i terorizam, , Institute for Serbian culture Pristina; Belgrade: Faculty for Business Studies and Law, University Union-Nikola Tesla, 2. revised edition.

[9] Tancic, D. (2012) Osnove istorijskog metoda, Institute for Serbian culture from Pristina, temporary seated in Leposavic, Belgrade

[10] Tanic, Ž. (1961) *Metod posmatranja (analize) dokumenata*, Institute of Social Sciences, Belgrade

[11]   Collection of works: 'Comparative Research Methods' edited by Donald P. Warvwick & Samuel Osherson, Prentice-Hall, Inc., Englewood Clifss, N.Y., 1973.

[12]   Dirkem, E. (1982) O podeli društvenog rada, Belgrade.

[13]   Enciklopedija Britanika, knjiga 3, sesto izdanje, politika- Narodna knjiga, Belgrade, 2005.

[14]   Lasswell, H. Lerner, D. de Pool. S. (1953) , *The Comparative Study of Symbols,* Stanford University Press

[15]   Lasswell, H. Leites, N. (1949) *Language of Politics,* Studies in Quantitative Semantics, New York

[16]   Ritter J. Grunder, K. (1989) Historisches Worterbuch der Philosophie, Band IX, s. 1314, Basel

[17]   von Bayme, K. (1974) Suvremene političke teorije, Zagreb

[18]   Baier, K. (1965) Morale Point of Veiw - A Rational Basis of Erics, New York

[19]   Krippendorf, K. (1981) *Content Analysis - An Intraduction to Its Methodology,* Sage Publications, Beverly Hills, London

[20]   Morgan, L. H. (1981) Drevno društvo, str. 315, Beograd

[21]   Nejgel, E. (1974) Struktura nauke, Nolit, Beograd

[22]   Milosavljevic,S. Radosavljevic, I. (2000) *Osnovi metodologije politickih nauka*, National Gazzette, Belgrade

[23]   Popper, K. (1974) Objective Knowledge, An Evolutionary Approach

[24]   Stojak, R. (1990) Metoda analize sadzaja, Institute for research of national relations, Sarajevo, DP 'Graficar', Tuzla

[25]   Adorno, T. (1962) *Protokolle und referate des Seminars - Probleme der qualitative Inhaltsanalyse,* Institute fur socialforschung der J. W. Goathe, Universitet, Frankfurt

[26]   Adorno, T. (1962) Einleitung in die Muziksociologie, Frankfurt

[27]   Mitchell, R. (2020) Improved analysis of fat content: machines that anasyze the fat content in grids are evolving into more potent processing instruments. *National Provisioner*, Vol. 234. Issue, 9.