



***Elżbieta Pieniążek-Niemczuk***

The University College of Applied Sciences in Chełm (Poland)

ORCID: 0000-0002-9761-3596

## **Application of Content Analysis in Research of Political Discourse**

### **Zastosowanie analizy treści w badaniach nad dyskursem politycznym**

#### **Abstract**

Content analysis is one of the most frequently applied methods for the research of various forms of discourse, including political discourse. As a research technique, it was developed in studies on communication and is now frequently used in linguistics, for example, in text analysis, systematic studies of written text or transcribed speech, as well as in research on nontextual message content.

The following article deals with the application of content analysis in research into political discourse and discusses the crucial concepts of the research method. Specifically, it describes the manifest and latent content of political texts and presents two notions particularly useful in the evaluation of content analysis, i.e., ‘reliability’ and ‘validity.’ Additionally, two broad approaches to content analysis are discussed. The first is qualitative content analysis, and the second is quantitative content analysis.

**Key words:** *political discourse, content analysis, qualitative content analysis, quantitative content analysis.*

## Abstrakt

Analiza treści jest jedną z najczęściej stosowanych metod w badaniu różnych form dyskursu, w tym dyskursu politycznego. Jako technika badawcza rozwinęła się w badaniach nad komunikacją i obecnie jest często wykorzystywana w językoznawstwie, m.in. w analizie tekstu, systematycznych badaniach tekstu pisanego czy transkrypcji mowy, a także w badaniach nad treścią przekazu nietekstowego.

Niniejszy artykuł porusza kwestię zastosowania analizy treści w badaniach nad dyskursem politycznym oraz omawia podstawowe pojęcia związane z tą metodą badawczą. Przedstawiono w nim pojęcia treści jawnej i ukrytej tekstów politycznych oraz omówiono dwa szczególnie przydatne w ocenie analizy treści terminy, tj. „rzetelność” i „ważność”. Dodatkowo omówiono dwa szerokie podejścia do analizy treści. Pierwsza to jakościowa analiza treści, druga to ilościowa analiza treści.

**Słowa kluczowe:** *dyskurs polityczny, analiza treści, jakościowa analiza treści, ilościowa analiza treści.*

## Introduction: Defining content analysis

Content analysis is one of the most frequently applied methods for the research of various forms of discourse, including political discourse. Nonetheless, researchers have not been unanimous in describing the crucial concepts of content analysis. As noted by Benoit (2011, p. 268), the work on elaborating the definition lasted for over half a century. However, it is Berelson's (1971, p. 18) definition of 'content analysis' that is the most frequently cited. It states that it "is a research technique for the objective, systematic, and quantitative description of the manifest content of communication."

Holsti's (1969) definition of the notion seems to be in alignment with Berelson's way of thinking about content analysis. According to Holsti, (Holsti, 1969, p. 14), "[c]ontent analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages." What can be inferred from the two definitions is that the core feature of content analysis is *objectivity*. However, this feature has been causing major disputes among researchers who, like Benoit, are far from calling it objective. Benoit argues that "practitioners" of the technique are just "human beings who attribute meaning to the numbers produced by this process" (Benoit, 2011,

p. 269), so objectivity should be seen as the final goal of content analysis rather than its intrinsic feature.

On the other hand, according to Krippendorff (2004, p. 18), “Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use.” The researcher makes reference to the issue of the context of content analysis and incorporates ‘validity’ as one of its crucial elements. However, this definition can be debated, as valid inferences that possess real defining impact, should be contrasted with invalid ones. Researchers, however, are far more eager to report valid content analysis rather than invalid inferences. The reason is obvious; presenting invalid conclusions can be seen as a failure in the initial assumption, the choice of source material, or the selected methods and techniques. Still, the outcome would be research of little, some would say, scientific value.

The definition of content analysis offered by Riffe, Lacy and Fico (2005, p. 25) draws upon many of the ideas mentioned previously. The authors also recognize inferences about the context and describe the technique as systematic and valid. They point to the importance of a factor that they refer to as a “statistical method” and notice that content analysis relies on inferential statistics, like an assessment of connotations or contrasts. Moreover, they mention that content analysis is suitable to “draw inferences about the sources who produced those messages or draw inferences about the reception of those messages by the audience.”

As inferences about a context can be derived, the act of production and reception of a certain message is a new and crucial concept to which Benoit also refers. The scholar emphasizes the conviction that content analysis is, in fact, the technique whose main goal – but what is crucial to mention is that it is not a feature – is to achieve objectivity, validity, and reliability. (Benoit, 2011, p. 270) states that “content analysis quantifies dimensions (variables) of content in message texts.” To prove his point, the author presents some results of the research he had previously conducted. This research dealt with the content analysis of American presidential debates. What was observed is that according to the “topic” criterion, 75 percent of the messages produced during debate statements were focused on policy issues, while 25 percent were focused on character. He also noted that research with the use of the technique dealt primarily with the elements of messages that were verbal, i.e., words and concepts expressed by means of words. This, however, focusing on arguments, claims, and themes expressed in words is a matter of scholars’ convenience as reception frequently takes place in verbal dimension. (Benoit, 2011, pp. 269-270)

A researcher who intends to content analyze a (political) text ought to decide on three crucial aspects, namely sampling, unitizing, and coding (Alonso, Volkens and Gómez, 2012, p. 13).

1. **Sampling** – this is a technique aimed at measuring the content of groups of messages. Many scholars, including Krippendorff (2004), Riffe, Lacy, and Fico (2005), and Benoit (2011), have mentioned the types of sampling in their works. Those most frequent types distinguished by Benoit are presented below (2011, p. 272):
  - *Census sample* –this includes all constituent representatives of a given population of texts (e.g., presidential nomination acceptance speeches). A drawback of the method is that there are few research works that contain a complete census sample due to the fact that populations of texts steadily or rapidly grow in number. The only opportunity to conduct a complete census is to choose a population of texts whose “production” has ceased to a terminal, definite end;
  - *Random sample* – in this method, not every member of a given population of texts undergoes analysis, but all constituent members have an equal chance of being included in the sample. The method is considered one of the most desirable approaches because random samples make a useful generalization of the population of texts;
  - *Convenience sample* –this consists of sample texts that are conveniently available to the researcher. The method has some drawbacks, however. One of them is that the conclusions obtained from this sample will be of little generalization value. In addition, what should be noted is that some kinds of political texts are extremely difficult or impossible to obtain for sampling (e.g., face-to-face discussions among citizens);
  - *Purposive sample* – these are texts representing a particular, defined purpose (e.g., newspapers with a national range, major broadcast networks).
2. **Unitizing** –this is the designation of a segment of the text to a given category. The segments most frequently defined in political science are single words, parts or entire sentences, as well as various items, for example, films, books, speech articles. (Alonso, Volkens and Gómez, 2012, p. 15)
3. **Coding** – conducting the actual analysis of content requires developing a codebook that should address key concepts as well as the context unit, i.e., the part of the text that is used to interpret a given coding unit. There should also be a description of the coding process, the necessary definitions, and coding rules. (Benoit, 2011, p. 272)

Put simply, content analysis involves sampling, unitizing, and coding. The steps that follow these are comparing the developed categories and drawing conclusions from the research on the ‘manifest’ or ‘latent’ content of a text, an issue that is addressed in the next section of the article.

## **The manifest and latent content of political texts**

Content analysis also offers the possibility to make a distinction between 'manifest' and 'latent' content of messages. (Riffe, Lacy and Fico, 2005) An example of manifest content in political discourse includes issues addressed in a political message like health care, foreign policy, economy, or education. Thus, manifest content can be defined as explicit, clear elements that are present on the surface of a text. They are directly identifiable and therefore appear to have little, if any, ambiguity. (Neuendorf and Kumar, 2016, p. 2) Contrary to manifest content, latent content includes implicit, connotative aspects of a text that are frequently implied. They require inference and judgment about the character of the content and can cause ambiguity; latent content represents deeper meaning that cannot be seen on the surface of a text. (Neuendorf and Kumar, 2016, p. 2) An example of latent content in a political message might be identified with emotion (e.g., fear, pride, hope) that is expressed in a message. Other examples of latent content are irony or satire that carry meaning opposite to the literal content of the message. Thus, as noted by Benoit (2011, p. 273), the difference between manifest and latent content can be presented as evaluation versus identification. Taking this into consideration, a conclusion can be drawn that the surface features of a text are more likely to be measured, thus their reliability is greater. However, examining latent content is also useful and important, though it is based more on interpretation than factual data. Also, as stressed by Benoit (2011, p. 273), "latent and manifest content are not a mutually exclusive dichotomy," so establishing a sharp contrast between the two frequently poses problems. What can be observed is that more content analyses are conducted in manifest content as achieving high reliability in the field of latent content is more difficult.

## **Validity and reliability**

According to Benoit (2011), as well as Alonso, Volkens and Gómez (2012), two concepts seem especially useful in the evaluation of content analysis in research. These are 'reliability' and 'validity'. Neuendorf (2002, p. 141) sees reliability as a critical component of content analysis, without which all the measures conducted within the method are useless. Unless the same results are obtained by two or more coders, no matter what method is applied and what the circumstances of its introduction are, the research method cannot be considered reliable. Reliability is confirmed by numbers. However, when two coders obtain different results while examining the same coding units, the recipient cannot consider the results of both coders reliable and is forced to choose among the available interpretations. (Benoit, 2011, p. 273) Still it is impossible to ensure complete reliability as the chance of committing an error, either serious or minor, is ever-present and all-encompassing. For this reason,

the factor that ought to be the researcher's target is consistency between measurements. The greater the reliability of the research, the more consistent the results from recurring analyses of the same sample scope.

Validity, on the other hand, is confirmed when the results obtained from the research can be validated by facts. What should be noted is that some studies that rely on the content analysis technique do not discuss the validity of the research subject. One of the reasons for this is the fact that this feature is difficult to measure. However, validity can be achieved by correlation, i.e., by analyzing the same content using different analytic procedures. A strong relationship revealed by correlated data can provide the required support to validate the conducted research.

Thus, it can be concluded that both reliability and validity are subject to evaluation. Validity is constrained by unreliability, i.e., low reliability reduces validity; however, obtaining reliability cannot guarantee research validity.

### **Approaches to content analysis**

Benoit (2015) distinguishes two broad approaches to content analysis. The first, qualitative content analysis, classifies texts into a set of categories; then, a text is rated according to the previously designed scale to establish its quality. The second, quantitative content analysis, focuses solely on rating. As described by Neuendorf and Kumar (2016, p. 2), the distinction between the two approaches to content analysis may be contested by judging "whether the constructs of interest are principally quantitative or qualitative in nature, and whether the measures of these constructs result in quantifications or more qualitative (either microscopic or holistic) descriptions of the messages."

Quantitative research is rooted in agricultural research (Wright, 1921), while the early qualitative work was conducted in anthropology and sociology (Vidich and Lyman, 1998). The aspect of the appropriateness of various approaches available nowadays to content analysis has led to what is sometimes referred to as "paradigm wars" (Kohlbacher, 2005) among researchers – supporters of quantitative and qualitative research.

### **Quantitative content analysis**

Quantitative content analysis remains the dominant method in examining political messages. (Neuendorf and Kumar, 2016, p. 4) According to McCombs (2004), most linguistic research works using quantitative content analysis deal with issues like education, crime, taxes, or employment. In such research works, the occurrence of defined units in a sample text is counted. Aspects such

as the ways of expression in a text, text context, power relations in a text, agent perspectives, etc., are not dealt with while using this method. (McCombs, 2004)

The research using quantitative content analysis is usually preceded by the process of establishing categories to measure the content of messages. As rightly noticed by Berelson (1971, p. 147), “[c]ontent analysis stands or falls by its categories” – meaning roughly that the ability to define meaningful and appropriate categories is crucial to this method. Once named, they should be examined on meeting three criteria:

1. exhaustiveness, i.e., including particularly important parts of the content;
2. exclusiveness as parts of texts ought to be included in only one category; and
3. relevance, i.e., designing categories for the purpose of the research.

The process of naming categories can be achieved either deductively or inductively. The deductive way assumes that the names of categories can be found in the relevant literature (Verser and Wicks, 2006) and preferably from theory, but if no such names can be taken from this source, a researcher may rely on previous research that has been conducted. Alternatively, a scholar may choose to incorporate self-developed categories in the research. This inductive approach may, for example, start from a preliminary reading of source materials to acquire an idea of the content of the text and then prepare a list of topics. Another way would be to incorporate a systematic approach to aid the process of generating categories for a given research. Benoit and McHale (2003) incorporated the method of constant comparison (Glaser and Strauss 1967) in their research. Their aim was to establish the categories naming candidates’ personal qualities in presidential television spots. The result was to obtain four general categories and terms to fit within each of these. However, as noticed by Benoit (2011, p. 271), the possibilities for category creation are vast. They may be ordinal, interval, ratio, etc. However, frequency data is the most common dimension in the research of content analysis.

Nonetheless, the method has been severely criticized by researchers, especially by those practicing the theory of hermeneutics and applying critical theory to their research (Bucher and Fritz, 1989, p. 145–149) Some of the main reasons for the critique are that the method is grounded in the behaviorist paradigm of science (Langer, 1997, p. 6). Very often, the critics of the method put forward the argument that by breaking the text into categories that can be counted, the analyst is, in fact, destroying the subject of their study because the analysis neglects the textual whole and/or internal relations between the categories. Thus, there is a real threat that important aspects will not be taken into consideration (Asp, 1986, p. 21) Yet, because researchers are still able to obtain certain representative results from the textual analysis, it remains one of the most common tools in analyzing political messages.



## Qualitative content analysis

Berelson's book *Content Analysis in Communication Research*, which was first published in 1952, outlined the methods and goals of quantitative content analysis. In general, this approach focused on the assessment of frequency analyses. Qualitative content analysis emerged both as an opposition to and a critique of quantitative content analysis.

One of the most ardent critics of the assumptions developed by Berelson was Kracauer (1952), who in his article "The challenge of qualitative content analysis" argued that quantitative research failed to acknowledge the particular quality of texts under investigation and that the context should not be neglected. Moreover, the author claimed that patterns in a text can be examined by showing various possibilities for interpreting them, as they possess multiple connotations. (Kracauer, 1952, p. 637f) Mayring (2000, p. 6) describes quantitative research as "a superficial analysis without respecting latent content and contexts, working with simplifying and distorting quantification." Ritsert agreed with Mayring in saying that qualitative analysis neglects latent structures in texts and highlighted additional aspects that according to him are not dealt with by such analyses. These aspects include the context of text components and distinctive individual cases – things that do not appear in the text. (Ritsert, 1972, p. 19–31)

Schreier (2012, p. 21) defined the features of qualitative content analysis. According to him, it is interpretive, naturalistic, situational, reflexive, possesses emergent flexibility, is inductive, is case-oriented, and emphasizes validity. By means of this method, researchers are able to identify the thematic structures that exist in a text. Categories are not imposed on a researcher but appear as a result of the researcher's close study of a text. Thus, it can be stated that the method aims to interpret symbolic structures of the social and cultural substance of a text. (Neuendorf and Kumar, 2016, p. 4) Also, qualitative content analysis should be seen as an evolving process because its aim, i.e., discovery assumes constant comparisons. Larsen (1991, p. 67) believes that the method is not an end in itself, while Denzin and Lincoln (2017, p. 8) stated that "qualitative research is many things to many people." This sentence stresses the essence of qualitative content analysis, i.e., highlighting the depth of the notion and its vagueness. That is the reason why a single concise definition of the term cannot be found. The word "qualitative" emphasizes making distinctions based on the qualities of entities and units under examination. Rubin and Rubin (1995, p. 31) noted that the qualitative approach recognizes that "meaning emerges through interaction and is not standardized from place to place or person to person." This statement carries serious implications as it acknowledges that objective knowledge does not exist.



## **Mixed methods**

An opportunity to reconcile the disputes among qualitative and quantitative supporters has led to mixed methods gaining popularity. Attempts have been made to combine the advantages of both approaches. (Jick, 1979, p. 602)

Mixed methods allow for the same aspects to be examined by means of different approaches combined. Thus, the data collected in this way undergoes multimethod examination. As stated by Gillham (2000, p. 13), “[d]ifferent methods have different strengths and weaknesses. If they agree, then we can be reasonably confident that we are getting the true picture.” The effectiveness of this claim is based, to a great extent, on the assumption that “weaknesses in every single method will be compensated by the counter-balancing strengths of another”. (Jick, 1979, p. 604)

The view is supported by Neuendorf and Kumar (2016), among others, who claim that quantitative and qualitative approaches to content analysis are complementary and may be useful, for example in examining political texts. Nonetheless, the differences between the methods should be mentioned at this point. The first difference is related to the sampling procedure, which in the case of quantitative content analysis is most frequently probability-based, while in the qualitative content analysis, it is purposive, defined by context, structure, process, and the form of a political message. (Neuendorf and Kumar, 2016, p. 5) When the research goals of the two methods are compared, it can be observed that the quantitative approach is focused on the type, character, and possible outcomes of a political message. The qualitative approach, on the other hand, is focused on discovery, by means of which social and political changes are brought about. (Neuendorf and Kumar, 2016, p. 5)

## **Conclusions**

To sum up, content analysis can help to interpret the language of politics that is complex and indefinite. Understanding which tools are preferred or required can aid ‘persuasion professionals,’ including politicians, in achieving their goals. For this reason, research on linguistic constructions should incorporate a mixed method of content analysis – a combination of qualitative and quantitative approaches – to obtain the best understanding of the use, meaning, and potential future application of these tools as politics has become a self-conscious and sophisticated enterprise.

## Bibliography

- Alonso, S., Volkens, A. Gómez, B. (2012). *Content-analyzing Political Texts. A Quantitative Approach*. Spain: Centro de Investigaciones Sociológicas.
- Asp, K. (1986). *Måktiga massmedier. Studier i politisk opinionsbildning*. Stockholm: Förlaget Akademilitteratur.
- Benoit, W. L. (2011). Content analysis in political communication. In: R. L. Holbert, E. P. Bucy (eds). *Sourcebook for Political Communication Research. Methods, Measures, and Analytical Techniques*. London, United Kingdom: Routledge.
- Benoit, W. L., McHale, J. P., Hansen, G. J., Pier, P. M., McGuire J. P. (2003). *Campaign 2000: A Functional Analysis of Presidential Campaign Discourse*. Lanham, MD, United States: Rowman & Littlefield.
- Berelson, B. (1971). *Content Analysis in Communication Research*. (Reprint) New York: Hafner.
- Bucher, H-J, Fritz, G. (1989). Sprachtheorie, Kommunikationsanalyse, Inhaltsanalyse. In: Baacke. Dieter & Kübler (eds). *Qualitative Medienforschung. Konzepte und Erfahrungen*. Tübingen. Max Niemeyer Verlag.
- Denzin, N. K., Lincoln. Y. S. (eds). (2017). *Handbook of Qualitative Research*. Thousand Oaks, United States: SAGE Publications Inc.
- Gillham, B. (2000). *Case Study Research Methods*. London: Continuum.
- Glaser, B. G., Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Holsti, R. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison Wesley.
- Jick, T. D. (1979). "Mixing qualitative and quantitative methods: Triangulation in action". In: *Administrative Science Quarterly*, 24 (December).
- Kohlbacher, F. (2005). "The Use of Qualitative Content Analysis in Case Study Research". In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol. 7(1), art. 21.
- Kracauer, S. (1952). "The challenge of qualitative content analysis". In: *Public Opinion Quarterly*, no 16(4).
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Langer, R. (1997). "The Concept Of Discourse in The Analysis of Complex Communication Events". In: *Institut for Interkulturel Kommunikation og Ledelse, IKL. Copenhagen Business School. Working Paper / Intercultural Communication and Management*, no. 26.

- Larsen, P. (1991). Qualitative content analysis. Between "Critical Theory" and Semiotics. In: H. Ronning, K. Lundby (eds). *Media and Communication. Readings in Methodology, History and Culture*. Oslo: Norwegian University Press.
- Mayring, P. (2000, June). *Qualitative Content Analysis*. [online 28 paragraphs] Qualitative-research.net. <http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm>, 28.01.2018.
- McCombs, M. (2004). *Setting the Agenda: The Mass Media and Public Opinion*. Cambridge: Polity.
- Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.
- Neuendorf, K. A., Kumar, A. (2016). Content analysis. In: G. Mazzoleni (ed.). *The International Encyclopedia of Political Communication*. John Wiley & Sons, Inc.
- Riffe, D., Lacy, S., Fico, F. (2005). *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Ritsert, J. (1972). *Inhaltsanalyse und Ideologiekritik. Ein Versuch über kritische Sozialforschung*. Frankfurt a.M.: Athenäum Fischer Taschenbuch Verlag.
- Rubin, H. J., Rubin, I. S. (1995). *Qualitative Interviewing, The Art Of Hearing Data*. Thousand Oaks: Sage.
- Verser, R., Wicks, R. H. (2006). „Managing voter impressions: The use of images on presidential candidate websites during the 2000 campaign”. In: *Journal of Communication*, no. 56.
- Wright, S. (1921). "Correlation and causation". In: *Journal of Agriculture Research*, no 20.0

*Correspondence concerning this paper should be addressed to*

**Elżbieta Pieniążek-Niemczuk**, Ph. D. in linguistics; Assistant Professor – a faculty member of the Department of English Studies at The University College of Applied Sciences in Chełm (Poland).

E-mail: [epieniazekniemczuk@nauka.panschelm.edu.pl](mailto:epieniazekniemczuk@nauka.panschelm.edu.pl)