# Content Analysis and the Social Sciences:
# Nature, features and uses

## Abdelhak ZIDANE[1], Ramdane MEHIRI[2]

*[1]Mohamed Khider University of Biskra, Algeria*
Abdelhakzidane@gmail.com
*[2] Mohamed Khider University of Biskra, Algeria*
ramdanemehiri@hotmail.fr

**ABSTRACT:** *Regarding the fact that there are many data analysis methods, this research concerns itself with content analysis which has been increasingly used in research recently. Furthermore, content analysis is seen as one of the main research techniques used to analyse data in the social sciences field. It is interested in analysing data within its specific context. This is, a data analysis method whose importance makes it useful for diferent academic desciplines such as sociology, cultural studies, philosophy, anthropology, literary studies, linguistics, etc. Moreover, it can be done both quantitatively and qualitatively although it has been originally used as a quantitative research method until recent decades. Broadly speaking, content analysis examines communication via texts or transcripts reaching the central aspect of social interaction, and providing insights into complex models of language use. Nevertheless, if we compare the growing body of research conducted on topics of the field of research methodology with research on data analysis, we find that the available literature devotes insufficient room for data analysis in general and content analysis in particular. Thence, the present study attempts to introduce it to students and uncover its merits focussing on its nature, features and uses.*

**الملخص:** بما انه يوجد العديد من طرق تحليل البيانات، فإن هذا البحث يهتم بتحليل المحتوى الذي اصبح استخدامه متزايدا في البحث مؤخرًا. زيادة على ذلك، فان تحليل المحتوى ينظر اليه كإحدى تقنيات البحث الرئيسية المستخدمة لتحليل البيانات في أبحاث العلوم الاجتماعية. حيث يهتم بتحليل البيانات في سياقها المحدد. مما يجعله مفيدًا لتخصصات أكاديمية مختلفة مثل علم الاجتماع ، والدراسات الثقافية ، والفلسفة ، والأنثروبولوجيا ، والدراسات الأدبية ، واللغويات. وإذا ما قارنا الكم المتزايد من الأبحاث التي أجريت حول مواضيع مجال منهجية البحث خاصة في تحليل البيانات ، نجد أن الأدبيات المتاحة لا تخصص مساحة كافية لتحليل البيانات بشكل عام وتحليل المحتوى بشكل خاص. ومن ثم تحاول الدراسة الحالية تعريفها للطلاب والكشف عن مزاياها مع التركيز على طبيعتها وميزاتها واستخداماتها.

**الكلمات المفتاحية:** تحليل البيانات، تحليل المحتوى ، البحث النوعي ، البحث الكمي ، التقنيات التمثيلية .

## 1. Introduction

Although data analysis in any research is a highly important step, doing research about it reveals that it is not given equivalent importance in the literature, comparing it to other parts of research methodology. That is to say, research on data analysis is somehow underrated. Therefore, this present research is concerned with data analysis, more particularly, content analysis. Furthermore, this latter is known for its interest in analysing all sorts of text. It, thus, pays special attention to the content of any textual material. It has its own methodology of analysing, discussing, and reporting data. So, in this paper, we attempt to explore this matter and uncover its merits. Noteworthily, this research relies heavily, but not exhaustively, on Klaus Krippendorff's book, which is entitled "*Content analysis an introduction to its methodology",* as a main reference for it is a thorough book about the targeted data analysis method.

## 1.1. Statement of the problem

During the last two decades, content analysis has been increasingly used in many fields of research. However, it can be noticed that it is not given much importance in the field of social sciences in general, and applied linguistics in specific. Perhaps this is due to students and novice researchers' unawareness of its extreme significance to deal with communications and texts.

## 1.2. Aims of the study

The major aim of this study is to stress on the importance of content analysis, and to reveal its benefits and merits for students and novice researchers. Secondarily, this is about a brief and easily accessible piece of information for them. In other words, this is a shortcut for knowledge about this type of data analysis. Finally, we reckon that there is no room for intuition in research, i.e., it has to be done systematically and scientifically, therefore, this research aims at showing the way content analysis is done methodologically and scientifically.

## 2. Conceptual Foundation

This part of the research introduces the conceptual framework of content analysis including its definition, history, and examples. Thus, this is to explain the basic principles of this data analysis method as to make the reader more familiar with it.

## 2.1 Definition

Generally speaking, content analysis is defined as a research method which employs certain procedures to "make valid inferences from text" (Weber, 1990, p. 117). That is to say, by analysing textual material, content analysis is used to make replicable and valid results. In addition, for Cohen et al. (2007), this research technique aims to summarise and interpret written data. Hence, it is known to be "a strict and systematic set of procedures for rigorous analysis, examination and verification of the contents of written data" (ibid, p.475). To explain, this method of analysis is for the purpose of dealing with that sort of data which is driven from texts in order to build reliable results.

Furthermore, content analysis allows the researcher to reduce large amount of written data to smaller chunks of sentences (Cohen et al, 2007).

Simply put, in the realm of this analysis technique, long texts can be reflected by a small number of expressions so that the analysis and discussion are easily conducted. Finally, a content analysis method can be used with any sort of written data, i.e., documents, questionnaires, interview transcripts, books, newspaper articles … etc

## 2.2 History

To give a thorough idea about this data analysis method, the following is about a brief historical development of it. According to Krippendorff (2004), systematic studies on texts can be traced back to the 17th century. Back then, it was mainly used to deal with religious texts, and after that it was developed through the years to reach its present version. Concisely, Margrit Schreier, in her book entitled *Qualitative Content Analysis in Practice*, gives a brief summary of the historical development of content analysis. For her, it can be divided into three major stages:

> Early applications, content analysis coming into its own, interdisciplinary and methodological elaborations. Early applications focused on the quantitative description and differentiation of newspaper content, often from a comparative perspective. The second phase was characterised by more sophisticated conceptualisation and measurement as well as an increasing interest in the effects of content on the recipients. During the third phase, content analysis came to be used in other social science disciplines. As the method was applied to novel kinds of research questions, even more variants were developed. This was accompanied by increasing attention to the context of production and reception and to the interrelation of selected textual characteristics (2012, p.13).

## 2.3 Manifest and latent content

Two forms of content are distinguished in the literature; manifest and latent content. The former is what is directly seen such as the words in an interview. Nevertheless, the latter refers to the underlying meaning of content such as the interpretation of that interview. The difference between these two forms of content is the objective or subjective nature of them. Frankly, the manifest content is the quantitative features of a text, but the

latent one is qualitative. Many content analysts include both forms as this gives a fuller picture of the phenomenon under study.

## 2.4 Components of Content Analysis

This part of the research concerns itself with the components of Content Analysis. These are the steps taken by the researcher in order to do this a data analysis method. Namely, Unitizing, sampling, recording, reducing data, inferring contextual phenomenon, and narrating the answer to the research question are the major components of content analysis.

### 2.4.1 Major Components

According to Krippendorff (2004, p. 83-86), unitizing is to systematically identify the text segments in which the analyst is interested. Simply, this is a process whereby the researcher makes a unit of the analysis. For instance, he may unitize the words, sentences, or paragraphs that he intends to analyse. Moreover, sampling, the second component, is an important step in this data analysis method. Usually, the content includes a large amount of data in content analysis, thus, it is difficult for the researcher to analyse the content of all units. That is to say, he has to use the technique of sampling in order to make the analysis less complicated.

Recording, or called also coding, is another Content Analysis component which serves as a way to get analysable representations. To explain, content analysis is, basically, a coding process. This latter is the process of transforming raw data into a standardized form. To put it differently, the researcher needs to classify what he observes into some quantifiable groups or into some standard category. Furthermore, next is the reducing data process. This is where the researcher is more likely to select only the needed data. For Miles and Huberman (1994): "data reduction refers to the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written up field notes or transcriptions" (p.10).

Inferring contextual phenomenon from texts is the fifth component of this data analysis method according to Krippendorff (2004). It "bridges the gap between descriptive accounts of texts and what they mean, refer to, entail, provoke, or cause" (p.85). This is to make relations between the mere words of the text and what effect they make in the real context.

Finally, the last component is, simply, to report data. Krippendorff (2004) calls it narrating the answer to the research question. Needless to say, this is about explaining the practical significance of the results or the contributions they make to the present literature.

## 2.4.2 Qualitative and Quantitative Content Analysis

As it can be noticed in the previous claims of this research method, two main approaches to content analysis exist: a qualitative and a quantitative one. Therefore, this part is interested in exploring each of the approaches under the realm of content analysis.

### 2.4.2.1 Quantitative content analysis

As it is mentioned earlier, content analysis in its beginnings used to deal only with quantitative approaches. Thus, quantitative content analysis identifies the frequency of thematic or rhetorical patterns and then discusses their relationship through inferential statistics. Moreover, quantitative content analysis is seen as a reliable method of analysis since it is based on objective perceptions. So, this approach to content analysis is more applicable by researchers who seeks to deal with features of textual material by systematically categorising and recording its parts so that they can be analysed and discussed.

### 2.4.2.2 Qualitative content analysis

For Mayring (2000), qualitative content analysis is defined as "an approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytic rules and step-by-step models, without rash quantification" (p.23). Simply put, qualitative content analysis permits the understanding of social reality in a subjective, yet scientific way; the exploring of meanings underlying physical messages.

## 2.5 Sampling

The available number of texts for a specific research problem might be too large to be examined as a whole, so content analysts need to limit their research to a manageable body of texts. Here, we are going to discuss sampling in relation to content analysis, and deal with some important sampling techniques that can be applied to texts.

### 2.5.1. Sampling Techniques Applicable to Texts

According to Krippendorff (2004), a sampling plan is needed when the researcher is dealing with a huge number of texts. That is, when the analysis of large amounts of textual material is not possible due to some time or financial constraints, sampling has to take place in order to make the content analysis feasible. Thus, in this part of the research we are exploring the most important, and used, sampling techniques in content analysis.

### 2.5.1.1. Relevance Sampling

Generally speaking, relevance sampling is a non-probability sampling technique that is selected based on characteristics of a population and the objective of the study. As far as content analysis is concerned, relevance sampling has the objective of selecting every textual unit that helps answering the research question (Krippendorff, 2004).

### 2.5.1.2. Convenience Sampling

Another sampling technique within content analysis is called convenience sampling. Broadly speaking, this sampling technique, also known as availability sampling, is a type of the non-probability sampling that is based on data collection from population members who are conveniently available to participate in study. Limiting ourselves to content analysis terminology, "A convenience sample is motivated by analytical interest in an available body of texts that is known not to include all texts of the population that the analysts are concerned with" (Krippendorff, 2004, P. 120).

### 2.5.1.3 Cluster sampling

With cluster sampling, the researcher divides the population into separate groups, called clusters. Then, a simple random sample of clusters is selected from the population. The researcher conducts his analysis on data from the sampled clusters. For Krippendorf (2004), at the level of content analysis, "Cluster sampling is the technique of choice when analysts cannot enumerate all units of analysis but find lists of larger groups of such units, or clusters" (p.116).

### 2.6 Analytical/ Representational Techniques

Krippendorf (2004, p. 191) has highlighted the roadmap of the analysis for the researcher. After coding texts and applying analytical constructs, the content analyst should rely on the following points:
- The inferences from text must be summarized to be easily understood and interpreted.
- The analyst has to discover patterns and relationships in the results so that to hypotheses concerning various relationships can be tested.
- The findings should be compared with results obtained by other researches to boost the validity of the content analysis at hand.

### 2.6.1 Tabulations

Tabulation is one tool of simplifying, comparing and investigating data. In content analysis, it is the most commonly used technique to make textual data understandable (Krippendorf, 2004). By definition, it is gathering similar recording units in pigeonholes and counting how many instances are found in each one. Tabulations are usually used to deal with frequencies. For instance, the number of words in each category in a text unit, percentages expressed relative to the sample size, proportions of a total, or probabilities can be dealt with in tabulations.

### 2.6.2 Cross-Tabulations, Associations, and Correlations

Cross-tabulation is another means used to categorise data. It is, needless to say, made of rows and columns defined by the categories classifying each variable. The content analyst is more likely to use it in order to look at the frequencies of "co-occurrences of values or categories rather than of simple categories" (Krippendorf, 2004, p. 195). This means that cross-tabulations are more complex than tabulations because they focus on associations and correlations between variables. They are generally used in statistical analysis to find patterns, trends, and probabilities within texts.

### 2.6.3 Multivariate Techniques

Multivariate analysis, in a broad sense, is a collection of methods that can be employed when various measurements are made on each individual or object in more than one sample. To put it differently, multivariate analyses are a set of techniques used to analyse units of data that include more than one variable, and the techniques are especially valuable when

working with correlated variables. As stated by Krippendorf (2004), multiple regression analysis is a widely used technique among content data analysts.

To explain, in a study "if there is only one dependent variable the statistical procedures are labelled univariate, regardless of how many independent variables there are. However, if a study has more than one dependent variable, then the procedures are referred to as multivariate" (Perry, F, 2017, p. 234). In other words, when the number of dependent variables is more than one, multivariate techniques have to take place. To illustrate, if a researcher is dealing with gender discrimination in newspapers, and the word "gender" incudes both males and females, which are two dependent variables, then he is doing a multivariate technique in his study.

## 3. Methods and Materials

For conducting the present research study, the researchers have adopted the descriptive method as it benefits the investigators to introduce the actual content of the research study in an easier and clearer manner to the master students and novice researchers. the descriptive method relies more on definitions, classifications, and details on the topic in question from different sources, and tries at the same to consider the subject matter from different perspectives to enables the student researchers to develop some cause-effect association among the study section variables. It offers guidelines to the aforementioned category of students, in this study, in examining the impact of Content Analysis on research in the social sciences. Content Analysis, receiving innovation and interest, also enables the new researchers in investigating the efficacy and effectiveness of many data collection methods (tools) and strategies.

## 4. Conclusion

As a recapitulation, this research dealt with the nature, features and uses of content analysis in order to reveal its significance in research for students and novice researcher. That is to say, it explored concepts, definitions, types, sampling techniques, and analytical techniques related to content analysis. The main objective of it was to boost the use of content data analysis in the social sciences in general and in applied linguistic research in particular. Its significance in analysing texts explains its growing use in many fields of research; this is why it has to be taken into consideration in our field of study.

# References

- Cohen, L., MANION, L., & MORRISON, K. (2007). Research methods in Education. New York: Routledge. [Electronic version]. Retrieved from http://knowledgeportal.pakteachers.org/sites/ knowledgeportal.pakteachers.org/file s/resources/ RESEARCH%20METHOD%20COHEN%20ok.pdf
- Krippendorff, K. (2004). Content analysis an introduction to its methodology. Beverly Hills: SAGE.
- Mayring, P. (2000). Qualitative content analysis. Forum: Qualitative Social Research, 1(2), 1-28. Retrieved from http://www.qualitative-research.net/index.php/ fqs/article/ view/1089
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage.
- Neuendorf, K. (2002). The Content Analysis Guidebook. Thousand Oaks: Sage Publications. [Electronic version]. Retrieved from http://depts.washington.edu/methods/readings/Neuendorf_2002.pdf
- Perry, F. L. (2017). Research in applied linguistics: Becoming a discerning consumer. London: Routledge.
- Schreier, M. (2012). *Qualitative Content Analysis in Practice.* London: SAGE
- Stemler, S. (2001). An overview of content analysis. [Electronic version]. Practical Assessment, Research & Evaluation, 7(17). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=17
- Weber, R.P. (1990). Content Analysis. [Electronic version]. Basic Content Analysis. Thousand Oaks: Sage Publications, 117-24. Retrieved from. http://wrt303sp08.files.wordpress.com/ 2008/01/ weber_content-analysis.pdf