DEBATES

Meeting the loose ends with ethnographic content analysis

Unindo as pontas soltas com a análise de conteúdo etnográfica

Paula do Espírito Santo
Isabel Soares

Abstract

Having the basis for a better, deeper and sustained discussion over a selected content and problem is one of the challenges one may face when a traditional research technique, such as content analysis, is applied in a less common approach as is ethnographic content analysis. The objective of this paper is to undertake a critical discussion about ethnographic content analysis, a specific approach that combines various technical resources, enabling new measures to achieving intriguing and rich results and in-depth discussion, considering a content corpus or data selected. This paper aims at contributing to a critical review of ethnographic content analysis, its advantages and limitations, as well as its potential as an alternative focus and strategy. The type of results expected from the implementation of this feature is either textual or statistical nature, which allows a more comprehensive approach and focuses on multiple details of the construction categorical process.

Keywords

Content Analysis; Ethnographic Content Analysis; Communication.

Resumo

Ter bases para uma melhor, mais profunda e sustentada discussão acerca de um conteúdo e problema selecionados será um dos desafios que se pode enfrentar quando uma técnica de pesquisa tradicional, como a análise de conteúdo, é aplicada num formato menos comum, como é o caso da análise de conteúdo etnográfica. O objetivo deste artigo é o de desenvolver uma discussão crítica acerca da análise de conteúdo etnográfica, uma abordagem específica que combina vários recursos técnicos, permitindo novas medidas para atingir resultados intrigantes e ricos e para uma discussão aprofundada, considerando um corpus ou material de análise selecionado. Este artigo pretende contribuir para fornecer uma visão crítica da análise de conteúdo etnográfica, as suas vantagens e limitações, assim como do seu potencial como um foco alternativo e estratégico. O tipo de resultados esperados desta implementação técnica é tanto textual como de natureza estatística, o que permite uma abordagem mais compreensiva e focada em múltiplos detalhes do processo de construção categórico.

Palavras-chave

Análise de Conteúdo; Análise de Conteúdo Etnográfica; Comunicação.

Old missions and different purposes for content analysis across history

To analyze, organize and understand further about the human communication content is an old objective and mission, which has been assuming whether different strategies and approaches as different are the purposes, reasons and means in its origin. Content analysis is, though, at a broader sense, such a natural process and technique as communication is inherent to the human behavior. The broader sense of content analysis is in its spirit, even when its application follows the most strict, formal and technologic way for developing new insights or organizing current tendencies. We enhance the importance of following the advantages of applying content analysis in a dual perspective, bearing on mind, on one hand a wide dimension and perspective that follows the communication process, in its inner original potentialities, as a broader reference, on one hand, considering whether the role of the researcher as the intentions of the producer of the content. And on another hand, complementing to this, it is expected throughout the application of content analysis to find and develop a concrete and focused path for following specific objectives, hypothesis and goals that allow us achieving findings with insights that can enable to get inner information and knowledge about selected communication contents. Content analysis has, though, two analytical conceptions, a broader, antique and deeper sense, which is as classic as critical understanding has been across the history of human knowledge, and a specific sense that should be found in a technique and analytic meaning, linked to, essentially, determined scientific resources or diversified disciplines. Since the beginning of the 20th century, the history of the development of content analysis made evident the benefit from the knowledge of various contributions, derived from different disciplines, each one, with its specific components, which contributed to an enriched and singular technique. The most prominent disciplines on the basis of content analysis technique raised its potentialities, in terms of enriching the findings achieved. Amongst these disciplines were mathematics, literature, communication studies, anthropology, statistics, political science, sociology, psychology, amongst the most evident. The development of content analysis was the result of several scientific disciplines and approaches, which contributed to turn this technique into one of the richest instruments, across the history of the methodology of social sciences, to organize and systematize information or data produced from the human communication process.

The history of the formal beginning of content analysis is rather atypical, considering the usual academic and scientific research based ways of introducing new research techniques in social sciences, and may explain its most frequent and current

contexts of application. In this case, the history of the development and validation of content analysis, we may consider, would turn into an interesting case of how can science and methodology be a direct consequence of the State and of the political power interest into it. In other words, the early development of content analysis, in the United States, its prior cradle as a technique, in modern times, received an important improvement, throughout the constitution of several 'experimental divisions', starting, mainly, during the World War II period. These research 'units', named frequently as Experimental Divisions, were a direct consequence of the will and were, mainly, maintained by the US Government, having an objective of achieving and decoding the symbolic and strategic meaning about military data. This data or information was, indeed, systematically produced by different media, specially, anchored in the press and on radio. Consequent to this, the mathematical basis of the communication theory built an important contribution to the content analysis development, essentially by the hand of Claude Shannon (1949), and his article entitled "Communication Theory of Secrecy Systems". A decisive impact had too the Shannon's and Warren Weaver (1949) popular book, *A Mathematical Theory of Communication*, which exposed in a mathematic and, simultaneously, simple basis, the effectiveness of the communication flow and how to solve its processing problems. Concerning to this, the interest about the communication processes and the secret side of information may be seen, though, as a secondary element of the history, in the film *A Beautiful Mind* (2001). This biographical drama based on the life of John Nash, a Nobel Laureate in Economics, contemplates the work of this researcher referred to the decoding process of press messages, with strategic and secret military interest, using supposedly mathematical decoding schemes, in a content analysis basis, although based in the cryptography principles. In fact, during the modern history, the Cold War period was one of the richest in what concerns the development of methods for producing and decoding strategic information, based in cryptographic principles. Diffie e Hellman would make, by the time, a significant contribution in cryptography, uniformizing the existent coding systems up to 1976. That would be the basis for the conception of one of the contemporaneous safest information system, the RSA algorithm, produced by MIT researchers ('RSA' as derived from the name of their researchers: Rivest; Shamir and Adleman).

Most of the standout social scientists, from the early 20th century, which contributed to the development of content analysis technique, such as Berelson, Lasswell, or Lazarsfeld, invested, simultaneously, in other research techniques, particularly in what concerns the polls surveys, producing fundamental knowledge in

what concerns methodology. These pioneer American contributions worked in several research fields, fundamentally in politics, propaganda materials particularly, communication, considering journalism as the most working area, and literature. These researchers contributed to a particular development of these areas and to the settlement of initial basis of research concerning the contents and method in these fields. In what concerns the communication and journalism, in particular, essentially, from the early thirties (LASSWELL, 1931; ALBIG, 1938; WAPLES, 1941; JANIS *et al.*, 1943; BERELSON and SALTER, 1946; BERELSON and GRAZIA, 1947). In 1952, Berelson would produce significant contributions, for settling the content analysis main principles and procedures (BERELSON, 1952b). From the earliest contributions we find an important need of validation, in terms of associating a quantitative and even formal basis in order to consider a consensual platform, so as to draw a recognized methodological basis for the acceptance of content analysis. This need for achieving a wide consent about this first validation technique step is understandable when revising the academic and scientific context for the acceptance of the several disciplines, sciences and techniques, at the beginning of the 20th century. For instance, in the twenties, the improvement in areas such as Psychology was, essentially, sustained and developed through the experimental method. Within the case of Psychology, behaviorism and the Pavlov experiences proved that the need for considering data and findings derived from an observable basis was a fundamental presupposition for getting the academic community acceptance. In other words, at the time, what was observable and quantifiable was likely to be considered scientific.

The inner conception about how to understand and apply content analysis was expressed in a clear perspective by Harold Lasswell *et al.* (1949, p. 40) considering "why, then, be quantitative about communication? Because of the scientific and policy gains that can come of it". Lasswell was responsible for significant work as director of the *Experimental Division for the Study of War Time Communications* (1941-1945). One of his works (LASSWELL, 1941), *The Garrison State*, would constitute a singular reflection about how propaganda was playing a fundamental role in a changing world where military power and violence were a basic structure to develop the political anchors of the State, with deep effective consequences, both at a symbolic and at a psychological level. This kind of concern was also present, in our view, considering Berelson's conception about the content analysis concept and its consequent data validity, giving origin to one of the most traditional views about this technique, reflecting, particularly, the importance of the specific political context of production. According to Bernard Berelson (1952a, p. 18), in *Content Analysis in Communication Research*, content analysis is "a research

technique for the objective, systematic, and quantitative description of manifest content of communications". This contribution opened the conceptual discussion basis to be followed in content analysis as an empirical technique with huge data resources, as far as the selection of the materials to be analyzed is concerned. In other words, though Berelson's definition was specific enough, considering the research practice campus, the ways to proceed the research, properly, were so wide as were the communication contents themselves (and will go on being). The deeper question about how to apply content analysis was much more directed on to which way to apply it? Would it be possible to include a qualitative approach for analyzing communication contents? The answer was yes, even in Berelson's point of view, despite its seminal definition. Content analysis was, indeed, a content research technique which was either quantitative or qualitative in its nature, though its origin, in conceptual terms, would enhance its "objective, systematic, and quantitative" features considering the "description of manifest content of communications". Following the same conceptual process of analysis concerning the technique, Cartwright (1953, p. 424) considered that content analysis refers to "the objective, systematic, and quantitative description of any symbolic behavior".

The inner understanding about which way to follow, considering the validity of the technique, was made by several contributors, accompanying the natural evolution of the technique' process of validation, through the history of social sciences. The sixties brought the beginning of a different conception about the potentialities of content analysis, in a qualitative perspective (STONE *et al.*, 1966; HOLSTI, 1969), developing the importance of assuming a new insight about content analysis, considering the role of the 'inferences' and of the 'context' as inner fundamental characteristics of this technique. A different perspective was being opened for designing and operationalizing content analysis as a qualitative research type technique, following a conception that enhanced, clearly, the importance of considering the development of 'inference' as a resource for deeper insights about the communication content. Following this perspective Stone *et al.* (1966, p. 5) in *The General Inquirer* defined content analysis as "any research technique for making inferences by systematically and objectively identifying specified characteristics within the context". This perspective was stressed and followed expressing conceptual evidence about how qualitative analysis should be compatible with 'objectivity' and 'regularity'. According to Holsti (1969, p. 14) "content analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages". Following an inferential perspective of content analysis application other contributions come just after (CARNEY, 1972; KRIPPENDORF,

1980b). Krippendorf (1980b, p. 21) defined content analysis as a "research technique for making replicable and valid inferences from data to their context". The data and the context would be the anchors for defining and developing the analysis process, fulfilling the major objective of allowing deeper and focused readings about the intentions of the producer as well as the inner complexity and idiosyncrasies about the messages produced. Robert Philip Weber (1990, p. 9), from Harvard University, defined content analysis as "a research method that uses a set of procedures to make valid inferences from text". Weber privileges the inferential perspective of content analysis, based on the message, the 'text', and, specifically, "about the sender(s) of the message, the message itself, or the audience of the message". The traditional organization of content analysis in its key elements, developed by Berelson, is followed here, confirming a convergence between an inferential and an 'objective' perspective of analysis, followed by the consequent evaluation of the causes, intentions and consequences of the message.

The conciliation of both the qualitative and quantitative perspectives is one of the most important steps made for a better use of the potentialities of content analysis. Andres Romero (1991), a researcher in the area of information theory and communication, from Complutense University of Madrid, would present an integrative definition of content analysis, directed to the analysis of communications contents. According to Romero (1991), "Content analysis consists in using methods, techniques and instruments, without distinction, which are used by researchers in information and communication analyze and explain objective, systematic, quantitative and qualitative forms and meanings of ideas, words, images current facts and that widespread and perpetrated by man can cause - and really cause - sociological and psychological reactions in the audience who are receiving those ideas, words, images and events, through which messages are disseminated by any means of social communication" (ROMERO, 1991, p. 15). This conceptual conciliation is representative of a new breath in what concerns the communications content study and allowed an integrative vision about how to analyze and interpret the wide richness of the several enriched corpus produced daily, in the several areas of communicational interest. The importance of analyzing the public impact of the messages is one of the characteristics that may count and be considered as a significant element for achieving an objective and substantive evaluation and interpretation of the considered contents of communication. To follow the consequences and impact of communications into the audiences is a measure that may introduces further elements for a better development of the original intentions and objectives of the message producer.

Berelson's concept of content analysis would be followed through the time, meaning its importance as a transversal and still valid perspective of how to understand the role and application of content analysis. Berelson's perspective is still present at Kimberly Neuendorf (2002, p. 1) definition of content analysis, enhancing that content analysis aims at the "*systematic, objective, quantitative analysis* of *message characteristics*". The importance of clarifying the scientific method in what concerns its application to content analysis is clear in this same contribution meaning that:

> [...] content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method, including attention to objectivity/intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing. It is not limited as to the type of messages that may be analyzed, nor as to the types of constructs that might be measured (NEUENDORF, 2002, p. 10).

According to Earl Babbie (1975, p. 348), in a classic definition, *content analysis is "the study of recorded human communications, such as books, websites, paintings and laws"*. These key applications and definition of content analysis were reinforced later, by Babbie, considering that "*content analysis* is a *social research* method appropriate for studying human communications through *social* artifacts" (BABBIE, 1975, p. 356; BABBIE, 2012, p. 188). These broad definitions allow a wider understanding of how content analysis may be a sustainable resource for a wide set of disciplines and contents, covering several types of messages, sources and purposes. In other words, the principal concern about content analysis and the principal tie that links together such different disciplines such as those that may included in the definition of 'recorded human communications' is the method and the capacity to find sustainable concepts to allow valid approaches and reliable analysis.

Conceptually, according to Espírito Santo (2010, p. 66) "content analysis is a technique which aims the systematization of information, according to the application of a coding, categorization and inference process, enabling a range analytical quantitative and/or inferential, depending on the objectives and analysis techniques". In a broader sense, content analysis is a technique that allows the identification, organization and interpretation of any communications' content, through a process of codification of the selected contents, considering a set of defined objectives or hypothesis and a specific context, through and supported in valid procedures of application, aiming to achieve reliable results.

Content analysis has made a long way and went through several concept revisions, through its history since the beginning of the 20th to the 21st century. Content analysis is, probably, one of the social sciences techniques that have still a rich way to follow considering its development as to follow patterns of validity. The human communications context, as an object, is the most challenging contents that can be found and, consequently, to find convergence and to settle defined anchors for applying its method of research is, probably, a never ending, scientific mission. This is why content analysis has to be evaluated by its potential to open new perspectives, new insights about the producers' intentions and the audiences' expectations, and less by the set of organized meanings that the message may allow. This is, for sure, the substantive and, simultaneously, symbolic meaning of choosing content analysis as a technique to improve social sciences research and to get closer to the meaning of human's communications.

The method and the work still in progress

Content analysis is a still dual technique, with two perspectives, quantitative and qualitative and inner problems of uniformization and, consequently, of reliability and validation. Reliability and validation are still opened problems to content analysis which resolution is far from having a defined, extensive and scientifically clear answer.

A critical perspective about the methodological fragility of content analysis has been enhanced, particularly when comparing content analysis to other quantitative techniques, in what concerns coding schemes. Several contributions (STEMLER, 2001; BRYMAN, 2001; KRIPPENDORF, 2004; RIFFE, LACY and FICO, 2005; NEUENDORF and SKALSKI, 2010) considered that the methodological consolidation of content analysis is still a work in progress, considering the wide range of methodological approaches and, particularly, of coding schemes, when considering the quantitative perspective of content analysis application. Reliability implies consistent classification procedures (KRIPPENDORF, 1980b; WEBER, 1990) and a clear discussion recording instructions in explicit terms (STEMLER, 2001, p. 2).

Reliability and validity are two integrative and co-related concepts, which are continuous. Different perspectives have been discussed, though at the fundamental the inner meaning may cover a circular essence. Krippendorf (1980a), following Campbell (1957) describes validity as "that quality which compels one to accept scientific results as evidence. Its closest relative is 'empirical truth'" and proceeds in distinguishing between:

- *Internal validity*: which is best designated by the term reliability;
- *External validity*: "validity proper".

According to Stemler (2001, p. 2) "reliability may be discussed in the following terms:

- *Stability*, or intra-rater reliability. Can the same coder get the same results try after try?
- *Reproducibility*, or inter-rater reliability. Do coding schemes lead to the same text being coded in the same category by different people?"

One of the classic solutions to measure reliability is based on a percent of agreement between raters, or the Cohen's Kappa method (COHEN, 1960). However, as stated, originally (COHEN, 1960), and, according to Stemler (2001, p. 2) "the problem with a percent agreement approach, however, is that it does not account for the fact that raters are expected to agree with each other a certain percentage of the time simply based on chance". This problem can be solved by the application of the following formula (COHEN, 1960):

$$K = \frac{PA - Pc}{1 - Pc}$$

Where:
PA: proportion of units on which raters agree
Pc: the proportion of units for which agreement is expected by chance

Table 1 – Example Agreement Matrix

		Rater 1			Marginal Totals
		Academic	Emotional	Physical	
Rater 2	Academic	.42 (.29)[*]	.10 (.21)	.05 (.07)	.57
	Emotional	.07 (.18)	.25 (.13)	.03 (.05)	.35
	Physical	.01 (.04)	.02 (.03)	.05 (.01)	.08
		.50	.37	.13	1.00

Note: *Values in parentheses represent the expected proportions on the basis of chance associations, i.e. the joint probabilities of the marginal proportions.
Source: Stemler (2001, p. 2).

Summing the values in diagonal we obtain the percent of agreement:

$$PA = .42 + .25 + .05 = .72$$

Multiplying the marginal values:

$$Pc = .29 + .13 + 0.1 = .43$$
$$Kappa = \frac{.72 - .43}{1 - .43} = .51$$

The final value means which proportion of agreement can be found between the raters. The decision about the reliability of the instrument depends on the proportion of agreement found. According to Stemler (2001), following Landis & Koch (1977) as well as Wheelock *et al.* (2000), the decision is positive after .60, which means that the level of agreement is significant enough to find develop the coding procedure.

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00- 0.20	Slight
0.21- 0.40	Fair
0.41- 0.60	Moderate
0.61- 0.80	Substantial
0.81- 1.00	Almost Perfect

Source: Stemler (2001).

A critical about the methodological fragility of content analysis has been enhanced, particularly when comparing content analysis to other quantitative

techniques, in what concerns coding schemes. Several contributions (KRIPPENDORF, 2004; RIFFE, LACY and FICO, 2005; NEUENDORF and SKALSKI, 2010) considered that the methodological consolidation of content analysis is still a work in progress, considering the wide range of methodological approaches and, particularly, of coding schemes, when considering the quantitative perspective of content analysis application. According to Neuendorf and Skalski (2010, p. 2) it is salient that there is still a lack of "reliability assessment" and that "a gap still exists between the rigor required for an acceptable (i.e., publishable) content analysis and that required for other quantitative methods such as survey or experimental techniques". One major recommendation is that the procedures should be clear, better explained, even uniformized, considering the wide range and dispersion of quantitative models, particularly those applications of two basic methodological constructs and analysis criteria, human coding and computer coding, which usually are applied in a complemented form.

Considering the need to find solid methodological basis for content analysis implementation, Neuendorf and Skalski (2010, p. 4) state that "a core recommendation, then, is that all content analyses should be guided by accepted methodological sources that are informed by an array of methodological and applied research experiences, and optimally not limited to a single discipline". The range of platforms is wide. The inner problem to this is that most of the computer assisted platforms do not have clearly defined research units, needing, though, further explanation in terms of codification and of the conceptual limits to each code adopted.

Table 1 and 2 present a set of CATA (computer-aided text analysis) computer programs and its correspondent types and validation criteria. These data show evidence whether about the dispersion of criteria as well as about the lack of justification, concerning these quantitative platforms, for applying content analysis.

Table 2 – Common CATA Computer Programs
Quantitative CATA Programs

Program	Author	Original Purpose
VBPro	M. Mark Miller	Newspaper articles
Yoshikoder	Will Lowe	Political documents
WordStat	Normand Peladeau	Part of SimStat, a statistical analysis package
General Inquirer	Philip Stone	General mainframe computer application (1960s)
Profiler Plus	Michael Young	Communications of world leaders
LIWC 2007	Pennebaker, Booth, & Francis	Linguistic characteristics & psychometrics
Diction 5. 0	Rod Hart	Political speech
PCAD 2000	Gottschalk & Bechtel	Psychiatric diagnoses
WORDLINK	James Danowski	Network analysis/communication
CATPAC	Joseph Woelfel	Consumer behavior/marketing

Source: Neuendorf and Skalski (2010, p. 31).

Table 3 – CATA Programs: Types and Validation
Quantitative CATA Programs

Program	Type	Validation
VBPro	Word count/researcher-created dictionaries only	N/A—all custom dictionaries
Yoshikoder	Word count/researcher- created dictionaries only	N/A—all custom dictionaries
WordStat	Word count/researcher created dictionaries only	N/A—all custom dictionaries
General Inquirer	Word count with pre-set dictionaries	No--Dictionaries adapted from Harvard IV, Lasswell values, other standard linguistic and socio-psychological scales
Profiler Plus	Word count with pre-set dictionaries	Proprietary
LIWC 2007	Word count with pre-set dictionaries (researcher created dictionaries may be added)	Some dimensions have been validated against assessments by human judges
Diction 5. 0	Word count with pre-set dictionaries	No—Based on R. Hart's substantive work
PCAD 2000	Word count with pre-set dictionaries (researcher created dictionaries may be added)	Long history of development of a human-coded scheme; both human & CATA heavily validated against clinical diagnoses
WORDLINK	Word co-occurrence	N/A—emergent dimensions
CATPAC	Word co-occurrence	N/A—emergent dimensions

Source: Neuendorf and Skalski (2010, p. 31).

Even if the CATA system is crucial for attending to urgent measures for reliable and valid analysis, according to Neuendorf and Skalski (2010, p. 3), the human coding techniques are also important as they are the support for understanding the logic of the CATA analytic schemes, attending that the human coding techniques are important for:

(a) the origination of content analytic schemes that eventually become CATA algorithms;
(b) the measurement of constructs that are highly latent (and, correspondingly, for which a line of research has not yet devised adequate CATA indicators);

(c) the ongoing validation of CATA measures, as language standards and practices (idioms, etc.) evolve over time.

Achieving reliable data is dependent, determinately, upon the human coding techniques, whether in quantitative as in qualitative analysis. The decisions made about the key concepts inclusion, its relevance and its role in the construction process have a structuring role for making clear and sustainable validity as a concept. Defending a naturalist or constructivist paradigm concerning the content analysis research, as a dynamic alternative to positivism, Erlandson *et al.* (1993) develop the perspective of working the validation process as a direct consequence and supported in the triangulation of procedures, including incorporating and discussing the role of different roles of data, procedures and theories. This conception privileges ethics and authenticity as the fundamental basis for achieving reliable data and to applying valid measures.

There are three requirements that sustain, in an integral vision, fidelity as an operative concept to content analysis (KRIPPENDORF, 1980b; WEBER, 1990). These principles converge between each other, and are simple and specific enough to be the prior guides to the mission that each researcher is obliged to follow.

- Stability;
- Reproduction;
- Relevance.

Stability is a quality that meets the permanency of the research development across the time, independent from the researcher objectives and intentions. Reproduction reflects the ability of developing reliable and generalizable data to be followed and replicable. Relevance is essential as a principle supported in the need to find and offer reference research lines consequent to each study developed. The application of content analysis implies, though, the evaluation of the production context as well as the each research object as well as the adequate and exhaustive definition of research objectives, in order to set a clear way to follow in methodological terms. The clear identification of a set of six key questions is, according to Krippendorff (1980b), a fundamental diagnosis to carry on in each content analysis study. These questions are the following:

1) Which data are analyzed?
2) How are they defined?

3) What is the population from which they are drawn?
4) What is the context relative to which the data are analyzed?
5) What are the boundaries of the analysis?
6) What is the target of the inferences?

The understanding and implementation of the method in content analysis has still in common the need to answer to the following primary questions:

7) What to look for?
8) Under what purposes?
9) Expecting to what?

According to Espírito Santo (2010, p. 69) "content analysis has general research objectives which are diversified and aim at the systematization and to finding tendencies, intentions, manifest or non manifest content, regularities, singularities of research, whether at the level of the style or at the level of the communicational contents". Considering those objectives, the context might be relevant considering objects where inference has a determinant role to get further and latent meaning and information. The success of the research might depend upon the following elements (ESPÍRITO SANTO, 2010, p. 69 e 70):

- The available resources in terms of time, human and material resources and technology. Regarding the latter remember the importance of access today, the computerized databases from various documents;
- The particularities of each type of material analysis;
- The knowledge and mastery of the techniques and procedures from the
  range of possibilities of content analysis research, the standpoint of data processing and results analysis used in each study;
- The idiosyncrasies of each researcher (or the designated 'equation staff ');
- The academic, political, social, cultural and economic context of the research.

The basic conceptual instruments of content analysis are:

- The categories;
- The indicators;
- The research units.

Validity must also be ensured be the categories identification and definition, as the key ideas of each research objective or hypothesis. The categories must be organized according to a set of principles that give them the necessary condition to sustain its validity. These principles are the following (BARDIN, 1977):

- Homogeneity (or being of an identical nature);
- Exhaustiveness (or contain all the key ideas or dimensions to the hypotheses or objectives referred);
- Exclusivity (or not to repeat or overlap the categories for the objectives defined);
- Objectivity (or, regardless of the researcher, the results must be the same);
- Adequacy or pertinence (or the appropriateness to the objectives, assumptions or contents analyzed).

These principles are also a guide to the following elements, which the indicators and the research units. The indicators are key-ideas derived from the categories. The research units must be defined according to the type of research, the context analyzed and the objectives enunciated for each research.

Defending objective and quantitative uniformized measures for settling the range of criteria applied is still a fundamental question, in terms of the content analysis technique validation. We defend the need to further convergence of measures and its criteria in order to develop valid, sustained analysis and the consequent discussion, in order to fulfill reliable assets and conclusions. To develop clear and convergent content analysis platforms is the major methodological challenge that content analysis faces and needs to solve, under the risk of taking huge steps without the proper or valid support for achieving the purposes meant for.

From qualitative to ethnographic: identity and boundaries

Content analysis has a unique method, independently of its different quantitative or qualitative approaches. To reduce information, aiming its comprehension and functionality, to set regularities, organizing information, according to objectives defined, sustained in a valid method and looking for reliable results are the key and elementary traits that guide theoretically the mission of the content analysis technique, since its origins.

Ethnographic content analysis is a type of qualitative content analysis, which relies in a particular relation of the analyst with the data and the process of working the contents selected. Observing closer qualitative and ethnographic content analysis we find an opened window to get improved achievements, derived not strictly from the data, but, mainly, from the context and from the capacity of the analyst to understand how to get closer, in inferential terms, in a valid way, with reliable coding schemes. The inner central question to this section is:

- How different qualitative and ECA are?

This is the key question which answer we look to define and discuss forward. The expression ethnographic content analysis (ECA) was, firstly, used by Altheide (1987) and had a purpose of giving identity to a concrete approach for developing the objective of organizing and formulating an inferential purpose of analysis, considering a specific corpus or data and a context of evaluation. ECA and qualitative content analysis are, in practical terms, two approaches, with the identical intentions or purposes, research process and *modus operandi*, however, diverging background theoretical identity, validity and reliability measures and type of results expected.

Table 4 – Comparison between qualitative analysis and ECA

	Qualitative analysis	ECA
Intentions or purposes	Comprehension of the meanings of communication; verification of theoretical relations	Comprehension of the meanings of communication; verification of theoretical relations
Background theoretical identity	Discipline and object centered	Grounded theory
Research process	Systematic, analytical, reflexive	Systematic, analytical, reflexive, not rigid
Modus operandi	Inference, inductive, comprehension	Inference, inductive, comprehension
Reliability	Description of data collection; verification and justification of the coding process; direct questions to the object; rather conciliation with other techniques (ex. interview)	Description of data collection; verification and justification of the coding process; direct questions to the context; frequent conciliation with other techniques (ex. interview)
Validation	Construction description of the empirical process, based on the object and context	Construction description of the empirical process, based on the object, and the researcher's position with the cultural context framing
Type of results expected	Context and object centered; new concepts	Researcher centered; new concepts

Source: Own elaboration.

Qualitative content analysis has an evident difference, when compared to quantitative approaches of content analysis, which is its angle of observation and, consequently, the starting basis to look for in terms of 'intentions' and research 'purposes'. Whether qualitative and ethnographic content analysis look for achieving further in terms of observation, behind the express meaning of the contents, through a comprehensive perspective, aiming to get inside deeper, concerning to the symbolic and latent meanings of communication. Regarding to this, Altheide (1996, p. 16) stresses that ECA is a technique which allows the comprehension of the meanings of communication as well as the verification of theoretical relations. The selection of objects usually determines the type of approach to be followed.

Considering the 'background theoretical identity', a critical review of the theoretical and methodological anchors ECA finds an evident, though nonlinear, path from ethnography to grounded theory as a prior theoretical basis, according to Altheide (1987). Originally defended by Barney Glaser and Anselm Strauss (1967), the grounded theory addresses the scientific purpose of generating and discovering

theory from data. According to Altheide, privileging collection of data which is the basic step for getting further and deeper sustained reflection and for allowing better and wider discussion about the concepts building and operating process. Though these similitudes between ECA and grounded theory, Altheide (1996, p. 17) defends that "grounded theory is trying to generate clear testable hypothesis as a foundation for 'theory'". ECA is not oriented to theory development but is more comfortable with clear descriptions and definitions compatible with materials. Central to both, however, is the importance of constant comparison, contrasts, and, theoretical sampling". The unique opportunity to work the meaning of significant data adds an essential role in QCA researches, which is intended to be used, as stated by Altheide (1987), "to verify or confirm hypothesized relationships, which enhances its empirical role for discovering new concepts and to sustain, even indirectly, theory". The grounded theory allows the construction of new concepts and is focused in getting theoretical improvement by the data reunited. According to Borgatti (1996):

> […] the grounded theory approach, particularly the way Strauss develops it, consists of a set of steps whose careful execution is thought to "guarantee" a good theory as the outcome. Strauss would say that the quality of a theory can be evaluated by the process by which a theory is constructed. (This contrast with the scientific perspective that how you generate a theory, whether through dreams, analogies or dumb luck, is irrelevant: the quality of a theory is determined by its ability to explain new data) (BORGATTI, 1996, p. 1).

To follow, describe and discuss the research process is a direct need of the option for researching aiming at any type of qualitative analysis, mainly, considering ECA. Regarding to the 'research process' it is, though, identical considering qualitative and ECA. According to Altheide, in EQA the purpose is being "systematic, analytical, not rigid" (1987, p. 68). This perception about the research process is identical in terms of the qualitative content analysis, in general. A singular characteristic of EQC might refer to its lack of rigidity, in terms of getting inner reflexive analysis, though, subject to a permanent revision and opened perspective of analysis, based in a comprehensive relation of the researcher to the context, culture, conjuncture of the object analyzed.

Regarding to the '*modus operandi*', the importance of inference is fundamental in ECA, by allowing the development of a personal interpretive process, which aims for the comprehensive analysis of the latent and symbolic meanings of each message. "The inference looks for the reading or the analysis beyond the results

developed from the prior systematization achieved with the use of categorization" (ESPÍRITO SANTO, 2010, p. 92). The inference process allows the permanent link between the data and its context, and consequently with the researcher's analyzing options. Stressing this idea, Krippendorf (1980b, p. 99) states that: "Inference allows the operationalization of interdependencies between data and context". Inference is not an accessory use for qualitative content analysis, since from its development depends, in a major extent, the application of the inductive method and of the inherent comprehensive approach.

'Reliability' is a challenging element about applying ECA, as a part of the validation process, as a whole. Its definition has the same critical characteristics that ethnography usually meets, which are directed to the consolidation of the analyst role, during the empirical and analysis process. In other words, besides the description of the data collection, the verification and justification of the coding process needs further attention, in what concerns the presence and relation of the analyst concerning to the research process. In the last section, the revision of the inner fundamental problems concerning reliability stressed the importance of focusing this process at each time, in every research, and with a particular attention, in what concerns a much more revisional perspective, such as the ECA approach. The verification and justification of the coding process is, though, a much unfinished process, regarding the ECA, consequent to a much inductive supported approach, enriched by the analyst eventual participation into the construction of the corpus context selected. The introduction of techniques such as the interview allow further strength in terms of understanding and sustaining the results anchored through the content analysis technique. The research questions are, consequently, directed to different angles of approach in content analysis, focusing the object or the context, according to a selection of a generally qualitative or an ECA approach, respectively.

The 'validity' regarding to EQA and qualitative research analysis have a common sharing of a set of key methodological compromises, experienced since the early XX century and, strictly, dependent upon reliability, as the 'internal validity' prior to any option of research, as mentioned before. The process of validation is, regarding to Altheide (1996) point of view, the major objective of EQA is the discovering process, which places particular emphasis on the validation process and less on building analytical fidelity. In other words, the inner problems and key aspects concerning the researcher's position to the object and the cultural context is, though, one of major elements to be focused upon and discussed, for achieving a clear application of the research assets of content analysis.

The 'type of results expected' from the implementation of the ECA is either textual or statistical nature (ALTHEIDE, 1996), which allows a, simultaneously, extensive and comprehensive approach, focused on multiple details of the construction of the categorical process, however with a researcher centered approach. Altheide (1987) starts from the ethnography concept, referring to the description of people and their culture (SCHWARTZ and JACOBS, 1979), to find a particular methodological orientation, the ECA, that provide, besides the current qualitative assets of the research analysis process, a participant perspective of content analysis. When compared to traditional content analysis, ECA is distinguished by the interactive role of the researcher, along the research analysis building process, including the choices about the data collection. ECA determines that type of training required of the researcher must be developed in terms of data collection. One of the enriching results of ECA is the production of new concepts which emerge during the investigation in a more reproductive manner than in conventional/quantitative and even in quality content analysis generally. Inference brings a particular, participative, perspective considering the position of the researcher along the analysis.

The end of the beginning in ECA: looking forward to reinforce the method

Two ways of approaching, linked together, considering the quantitative and qualitative sides of content analysis, allowing having a complementary and unique methodological basis for better advance in content analysis. The researcher has an inner responsibility of clarifying his methodological options, his data access, and though of explaining and justifying how his categorical determinants were built as well as the several concepts and linkages produced. ECA is a still singular approach, allowing the construction of a unique set of outputs, focused rather on the contents itself than in the research options and limits, context or production characteristics and in the researcher, as an effective variable that links together the data and the significance of the data achieved. ECA is a paradoxal approach in content analysis. ECA have the paradox of being formally developed as an identified technique and approach by the late eighties, by Altheide, however, being, simultaneously, the most traditional approach ever applied since the origins of content analysis, at the beginning of the 20th century, as a technique. In other words, the history of content analysis has been made since the beginning as an alternative resource, meant to allow getting further insights whether to the manifest as to the latent message contents.

ECA implies a constant exercise for allowing the conciliation of methodology, doctrine, experience, sagacity and creativity. ECA implies a supplementary effort in

order to allow a permanent ability for revising and adjusting the categories made, in order to build a valid understanding between the focus of the research and the environment that sustains the data collected. In other words, the principal challenge to ECA is the work about the reliability of data and the comprehension of how to incorporate, in a convinced and clear mode, the enriched role that the analyst plays, in order to get further intelligible and meaningful results. The application of ECA is a permanent challenge, based in an orthodox technical basis, which is made in an opened and deeper perspective, only making sense when all the empiric process of data discovering and concepts construction is made clear, through along the way made. The usual richness of the materials selected implies a constant effort to a clear and sustainable distinction between what's important and what's secondary, according to the objectives and intentions made previously. The major challenge here is that at each analysis step the researcher has to follow and understand the production context, however, the researcher himself is also a part of the context. That's why ECA allows deeper and enriched, though, permanently, unfinished results.

■ **Paula do Espírito Santo** é Doutora em Ciências Sociais. Atualmente trabalha no Instituto Superior de Ciências Sociais e Políticas (ISCSP) e Centro de Administração e Políticas Públicas - CAPP/Universidade de Lisboa – UL em Portugal. E-mail: paulaes@iscsp.ulisboa.pt.

■ **Isabel Soares** é Doutora em Estudos Anglo-Portugueses (Imperialismo e Jornalismo Literário). Atualmente trabalha no Instituto Superior de Ciências Sociais e Políticas (ISCSP) e Centro de Administração e Políticas Públicas – CAPP/Universidade de Lisboa – UL em Portugal. E-mail: isoares@iscsp.ulisboa.pt.

References

ALBIG, William. The Content of Radio Programs, 1925-1935. *Social Forces*, v. 16, n. 3, p. 338-49, mar. 1938.

ALTHEIDE, David L. Ethnographic Content Analysis. *Qualitative Sociology*, v. 10, n. 1, p. 65-77, Spring 1987.

ALTHEIDE, David L. *Qualitative Media Analysis*. Newbury Park, USA: Sage, *1996*.

BABBIE, Earl. *The Practice of Social Research.* Belmont, USA: Wadsworth Publishing Co., 1975.

BABBIE, Earl. *Social Research Counts.* Belmont, USA: Wadsworth, 2012.

BARDIN, Laurence. *L'analyse de Contenu.* Paris: PUF, 1977.

BERELSON, Bernard. *Content Analysis in Communication Research.* New York: The Free Press, 1952a.

BERELSON, Bernard. Democratic Theory and Public Opinion. *Public Opinion Quarterly,* v. 16, n. 3, p. 313-30, autumn 1952b.

BERELSON, Bernard; SALTER, Patricia J. Majority and Minority Americans: An Analysis of Magazine Fiction. *Public Opinion Quarterly,* v. 10, n. 2, p. 168-190, Summer 1946.

BERELSON, Bernard, GRAZIA, Sebastian de. Detecting Collaboration in Propaganda. *Public Opinion Quarterly,* v. 11, n. 2, p. 244-53, Summer 1947.

BORGATTI, Steve. *Introduction to Grounded Theory.* 1996. Paper accessed in: <http://www.analytictech.com/mb870/introtoGT.htm>. Accessed in November 2014.

BRYMAN, *Alan. Social Research Methods. Oxford: Oxford University Press, 2001.*

CAMPBELL, Donald T. Factors Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin,* v. 54, n. 4, p. 297-311, 1957.

CARNEY, Thomas. *Content Analysis:* A Technique for Systematic Inference from Communications. Winnipeg: University of Manitoba Press, 1972.

CARTWRIGHT, Dorwin P. Analysis of Qualitative Material. In: FESTINGER, Leon; KATZ, Daniel (Eds.). *Research Methods in the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1953. p. 421-470.

COHEN, Jacob. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* v. 20, n. 1, p. 37-46, Apr. 1960.

ERLANDSON, David A. *et al. Doing Naturalistic Inquiry:* A Guide to Methods. Newbury Park, USA: Sage, *1993.*

ESPÍRITO SANTO, Paula. *Introdução à Metodologia das Ciências Sociais.* Lisboa: Sílabo, 2010.

GLASER, Barney; STRAUSS, Anselm. *The Discovery of Grounded Theory:* Strategies for Qualitative Research. Chicago, USA: Aldine Publishing Company, 1967.

HOLSTI, Ole R. *Content Analysis for the Social Sciences and Humanities.* Reading, USA: Addison-Wesley, 1969.

JANIS, Irving L.; FADNER, Raymond H., JANOWITZ, Morris. The Reliability of a Content Analysis Technique. *Public Opinion Quarterly,* v. 7, n. 2, p. 293-296, Summer 1943.

KRIPPENDORFF, Klaus. Validity in Content Analysis. *Scholarly Commons,* 1980a. Paper accessed in: <http://repository.upenn.edu/cgi/viewcontent.cgi?article=1299&context=asc_paper>. Accessed in November 2014.

KRIPPENDORFF, Klaus. *Content analysis:* An introduction to its methodology. Beverly Hills, USA: Sage, 1980b.

KRIPPENDORFF, Klaus. *Content analysis:* An introduction to its methodology. Thousand Oaks, USA: Sage, 2004.

LANDIS, J. Richard; KOCH, Gary G. The measurement of observer agreement for categorical data. *Biometrics,* v. 33, n. 1, p. 159-174, Mar. 1977.

LASSWELL, Harold D. The Measurement of Public Opinion. *American Political Science Review,* v. 25, n. 2, p. 311-326, May 1931.

LASSWELL, Harold D. *The Garrison State*. *The American Journal of Sociology*, v. 46, n. 4, p. 455-468, Jan. 1941.

LASSWELL, Harold *et al*. *Language of Politics*: Studies in Quantitative Semantics. New York: McGrawHill, 1949.

NEUENDORF, Kimberly A.. *The Content Analysis Guidebook*. Thousand Oaks, USA: Sage Publications, 2002.

NEUENDORF, Kimberly A.; SKALSKI, Paul D. *Social Science and Social Computing Workshop*, University of Hawaii, Honolulu, USA, May 23, 2010.

RIFFE, Daniel; LACY, Stephen; FICO, Frederick G. *Analyzing Media Messages*: Using Quantitative Content Analysis in Research. New York: Routledge, 2005.

ROMERO, *Andrés*. *Metodologia da* Análise de Conteúdo. Lisboa: Universidade Católica Portuguesa, 1991.

SCHWARTZ, Howard; JACOBS, Jerry. *Qualitative Sociology*. New York: Free Press, 1979.

SHANNON, Claude. *Communication Theory of Secrecy Systems*. 1949. Paper accessed in: <http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf>. Accessed in November 2014.

SHANNON, Claude; *WEAVER, Warren. The* Mathematical Theory of Communication. Illinois, USA: Illini *Books, 1949.*

STEMLER, Steve. An Overview of Content Analysis. *Practical Assessment, Research & Evaluation*, v. 7, n. 17, 2001. Paper acessed in: <http://pareonline.net/getvn.asp?v=7&n=17>. Accessed in November 2014.

STONE, Philip J. *et al*. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press, 1966.

WAPLES, Douglas. Press, Radio and Film in the National Emergency. *Public Opinion Quarterly*, v. 5, n. 3, p. 359-382, 1941.

WEBER, Robert Philip. *Basic Content Analysis*. California: Sage Publications, 1990.

WHEELOCK, Anne; HANEY, Walt; BEBELL, Damian. What can student drawings tell us about high-stakes testing in Massachusetts? *Teacher College Record*, 2000. Paper accessed in: <http://www.tcrecord.org/Content.asp?ContentID=10634>. Accessed in November 2014.