
Valeriano Piñeiro-Naval
http://orcid.org/0000-0001-9521-3364
vale.naval@labcom.ubi.pt
Universidade da Beira Interior

Submitted
April 15th, 2019
Approved
January 24th, 2020

© 2020
Communication & Society
ISSN 0214-0039
E ISSN 2386-7876
doi: 10.15581/003.33.3.1-15
www.communication-society.com

2020 – Vol. 33(3)
pp. 1-15

How to cite this article:
Piñeiro-Naval, V. (2020). The
content analysis methodology.
Uses and applications in
communication research on
Spanish-speaking countries.
Communication & Society, 33(3), 1-
15.

The content analysis methodology. Uses and applications in communication research on Spanish-speaking countries

Abstract

Based on a conceptual review of the content analysis research method, this article identifies meta-research studies that confirm the relevance of this technique in the atomised field of communication sciences. In light of the empirical evidence, this study is based on an analysis of 262 papers that use this tool and were published in major Spanish-language journals between 2013 and 2017. At instrumental level, bibliometric exploration is combined with content analysis itself, considering 18 categories, such as authorship, funding, sampling method and reporting of intercoder reliability. The results indicate that there is a steady annual growth in the number of papers produced by researchers affiliated to Spanish universities. These papers have more than two authors on average and often lack additional financial resources. As for the implementation of the content analysis method, it is generally applied to non-probabilistic samples and does not involve any intercoder reliability assessment and reporting. A series of recommendations are offered for researchers in Spain and in Latin America to meet international standards.

Keywords

Content analysis, research methods, meta-research, journals, communication, Spain, Latin America.

1. Introduction

This initial section compiles some of the paradigmatic definitions of content analysis, highlighting its main features and the stages that its proper execution requires. Due to length limitations, this theoretical review will be brief.

From a chronological point of view, it is necessary to go back to the 1950s, when Berelson (1952) laid the theoretical foundations for “a research technique for the objective, systematic and quantitative description of the manifest content of communication, for the purpose of interpreting it” (p. 18)¹. For his part, Holsti (1968) points out that content analysis can be used for “making inferences by systematically and objectively identifying specific characteristics of messages” (p. 601). Both authors –pioneers in the area– agree that content analysis enables a rigorous examination of information, i.e., the central element in the classic communication model: sender, message, receiver.

¹ However, content analysis had been already used in the 1940s by Paul F. Lazarsfeld, who led numerous empirical studies of this type at the “Bureau of Applied Social Research” at Columbia University (Martín-López, 1963).

Two decades later, Bardin (1986) expanded the definition of content analysis by pointing out that it is a technique designed to “obtain, by means of systematic and objective procedures of the content description of the messages, indicators (quantitative or not) which enable the inference of knowledge regarding the production and reception conditions” (p. 29). Therefore, an analysis of the information allows researchers to infer aspects related to its production and subsequent decoding, which are the other two elements in the traditional communication model. As Krippendorff (1990) summarises, content analysis is “a technique aimed at formulating, on the basis of certain data, reproducible and valid inferences that can be applied to their context” (p. 28).

In another order of ideas, Downe-Wamboldt (1992) remarks that content analysis is a means to describe and quantify specific phenomena based on oral, visual, or written parameters. Thus, content analysis is not limited to messages, but also considers the container, addressing the meanings as well as the signifiers of communication (Bardin, 1986). Regarding this disjuncture between “what” versus “how,” Naccarato and Neuendorf (1998) have recognised the differences that exist between “formal variables,” which are linked to the characteristics of the medium and cannot be transferred to other media types, and “content variables,” which exist regardless of the medium. In other words, despite its name, content analysis can be used to examine both formal and textual elements.

To conclude this conceptual review, Lombard, Snyder-Duch and Campanella (2002) emphasise that content analysis “is specifically appropriate and necessary for (arguably) the central work of communication scholars, in particular those who study mass communication: the analysis of messages” (p. 587). Comparing this social research tool with a microscope, a common device used in the natural sciences, Igartua (2006) explains that content analysis “allows us to discover the DNA of media messages” (p. 181). Therefore, and beyond its inferential and contextual capacity, it seems that the information transmitted by the media is the preferred object of study of content analysis.

On the other hand, as mentioned above, three adjectives are used to describe the fundamental features of content analysis: systematic, objective, and quantitative (Wimmer & Dominick, 2011). According to Lozano (1994), content analysis is systematic because it “is based on a set of procedures that apply in the same way to all the analysable content” (p. 142). The second feature, objectivity, implies that the researcher’s idiosyncrasies or biases do not affect the results, so that “when different people apply the same categories separately to the same sample of messages, they can reach the same conclusions” (Lozano, 1994, p. 142). These two features raise no doubts among the scientific community, unlike the quantitative condition of content analysis. In this regard, Piñuel (2002) states that content analysis relies on “measurement techniques, which are sometimes quantitative, like statistics based on unit count, and sometimes are qualitative, like rationales based on the combination of categories” (p. 2), and that both have the basic purpose of understanding the structure of texts. In order to delve a little further into this matter, it is necessary to take into consideration the following reflection by Espín (2002):

Quantification has been accepted as an important feature of content analysis. However, there are differences about what is meant by “quantitative.” For some quantitative researchers, it is synonymous with numbers; while others talk about a quantification degree, so that elements are described with such terms as “more,” “less,” “increment,” etc. It is therefore possible to discuss not a quantitative-qualitative dichotomy, but rather a continuum that ranges from the mere presence-absence of the attribute measured in a text, to the frequency and even intensity with which it appears (Espín, 2002, p. 97).

In this article, it is understood that content analysis is quantitative because, as Wimmer and Dominick (2011) clarify, its purpose is to achieve as precise a representation as possible of a wide range of messages. This representation is achieved by assigning numbers –codes– to the different categories adopted by the study variables, in such a way that they are quantified to

be processed statistically. This does not mean that all dimensions of analysis for messages (or variables) must be quantitative, since those governed by nominal (dichotomous or polychotomous) and ordinal measurement scales are very common in content analysis research, and both are qualitative in nature. On the other hand, strictly speaking, quantitative variables (interval or ratio) generally have a lower presence in codebooks.

Finally, with regards to the steps required to perform content analysis, experts in this field (Bos & Tarnai, 1999; Neuendorf, 2017; Riffe, Lacy & Fico, 2014; Weber, 1994; White & Marsh, 2006; Zamith & Lewis, 2015) broadly agreed on the following:

1. Establishment of objectives
2. Formulation of research questions and/or hypotheses
3. Identification of the research population
4. Selection of the study sample
5. Design and validation of the codebook and the coding form
6. Training of coders
7. Coding of sample using the designed tools
8. Assessment of the objectivity of analysis: intercoder reliability
9. Statistical analysis of data
10. Drafting of final research report

It is a logical and staggered formula, so it is highly recommended to follow it sequentially. To use content analysis effectively, it is necessary to establish, from the very beginning, what the main research interests are since, as Deacon et al. argue (1999), it is an extremely direct method: it gives answers to questions that are posed. The following section puts the emphasis –also at theoretical level– on an aspect of particular relevance within the whole process: intercoder reliability assessment.

1.1. *Reliability of the coding plan*

As Krippendorff (2011) points out, “reliability is the extent to which different methods, research results, or people arrive at the same interpretations or facts” (p. 94). More specifically, “it is expressed as a function of the agreement reached among coders on the allocation of units to the various categories” (Krippendorff, 1990, p. 197).

In general terms, three techniques are used for reliability assessment: simple agreement, chance-corrected agreement, and covariation (Neuendorf, 2017). Simple agreement refers to the percentage of agreement of coded units between two independent coders, i.e., it counts the number of ratings in agreement or disagreement. Chance-corrected agreement assumes that some of correspondence between coders is due to chance and corrects for this. Covariation, in short, measures whether the scores assigned by coders go up or down together, but not necessarily in precise agreement. Neuendorf (2017) argues that it is not acceptable to assess reliability without correction for chance, although simple agreement can be reported as a heuristic element along with another type of test. Table 1 presents a selection of the most common indicators:

Table 1: Indicators used to measure intercoder reliability.

Type	Name	Value Ratio	Desirable Values	Description
Index to measure simple agreement	Proportion Agreement Observed (PA_o)	0% (no agreement) 100% (total agreement)	~80%	Intuitive and used only for qualitative variables and for 2 coders
Index to measure chance-corrected agreement	Cohen's $Kappa$ (κ_c)	-1 (total disagreement) 1 (total agreement)	≥ 0.70 (exploratory studies: ~0.60)	Suitable for qualitative variables and 2 coders
	Krippendorff's $Alpha$ (α_k)	0 (no agreement) 1 (total agreement)	≥ 0.80 (exploratory studies: ~0.70)	Suitable for all types of variables (qualitative and quantitative) and ≥ 2 coders
Index to measure covariation	Pearson Correlation (r_p)	-1 (total disagreement) 1 (total agreement)	≥ 0.30 : substantial ≥ 0.50 : high	Used for quantitative variables and 2 coders

Source: Own elaboration, using data from Hayes & Krippendorff (2007), Igartua (2006), Krippendorff (2017), and Lombard, Snyder-Duch & Campanella (2002).

Due to the diversity of options –a total of 22, according to Zhao, Liu & Deng (2013)– there is some debate about which is the most appropriate index. Following Hayes and Krippendorff (2007), it is estimated that Krippendorff's Alpha is the most suitable since “it generalizes across scales of measurement; can be used with any number of observers, with or without missing data; and it satisfies all of the important criteria for a good measure of reliability” (p. 78). This is certainly the most critical phase of the whole system, because it will show the objectivity and transparency of the results (Piñeiro-Naval *et al.*, 2018).

1.2. Empirical research and the prominence of content analysis

Before examining the main empirical research studies that have tried to measure the prominence of content analysis in communication research, it is necessary to dwell on one aspect that is common to all of them: meta-research, a modality of research that “makes it possible to quantitatively describe how researchers develop their research practices, how scientific knowledge is disseminated and the use of theories in current communication research” (Saperas & Carrasco-Campos, 2019, p. 227). Thus, its objective is to analyse research production itself, which is very useful and necessary to measure scientific and, consequently, human progress (Ioannidis, 2018). In essence, it serves to map authors, topics, objectives and research methods (Caffarel-Serra, 2018), and helps us understand how a specific area of knowledge evolves; in this case, the area of social and media communication, studied under the methodological prism of content analysis.

In Berger's view (2016), content analysis is “one of the most commonly used research methodologies among scholars dealing with media and communication” (p. 390). In fact, a search for the English keyword “content analysis” provides over 2.2 million results. On the other hand, a search for the Spanish term “*análisis de contenido*” in Google Scholar yields about 107,000 results². In any case, both figures confirm the relevance of this method, which is the protagonist in the works cited below.

Starting with Spain, Bermejo-Berros (2014) has analysed the 339 papers published by *Revista Latina de Comunicación Social* in the 2004–2013 period. According to the author, 146 papers adopt a positivist approach, and together use 11 different types of research techniques. Content analysis accounts for one-third of all research techniques and the press is the

² Retrieved from <https://scholar.google.es> (04.12.2019).

preferred object of study. On the other hand, of the 353 authors identified, 257 are Spanish, 27 Mexican and 17 Argentinian. These three nationalities are the most represented in the sample. In addition, the disciplines that bring together the greatest number of authors are journalism and audiovisual communication, followed some way behind by advertising, public relations, and other areas.

A study by Goyanes, Rodríguez-Gómez and Rosique-Cedillo (2018) seeks to shed light on the state of research in communication sciences based on an analysis of the 3,653 papers published in the main Spanish journals between 2005 and 2015. One of their regression analyses establishes a positive association between maturity of the discipline and likelihood of producing empirical research based on content analysis. Similarly, although these authors eventually refuted it, they hypothesised that unfunded research is more likely to use content analysis than funded research, which is a circumstance that will be discussed later.

To conclude this brief tour through the Spanish context, it is important to review the main findings of the latest study on use of content analysis carried out by Martínez-Nicolás, Saperas and Carrasco-Campos (2019). Their study focuses on a sample of 1,098 works published between 1990 and 2014 by six Spanish journals: *Anàlisi*-Universidad Autònoma de Barcelona, *Comunicación y Sociedad*-Universidad de Navarra, *Estudios sobre el Mensaje Periodístico*-Universidad Complutense de Madrid, *Zer*-Universidad del País Vasco, *Comunicar*-Grupo Comunicar, and *Revista Latina de Comunicación Social*-Universidad de La Laguna. The study shows that, from 2000 onwards, the study of content –message– reaches a position of absolute prominence –in 50% of the papers– which implies that the most frequently used technique is content analysis: representing one third of all papers in the entire sample and 40% when focusing on the 2010–2014 period.

At international level, the classic work of Riffe and Freitag (1997) is a compulsory reference. They examined 486 papers that used content analysis and were published in the 1971–1995 period by *Journalism & Mass Communication Quarterly*. The sample represents 24.6% of all papers published by that journal in that period ($N = 1,977$). Among the most notable findings, Riffe and Freitag suggest that 46.7% of papers focus on the press and 24.3% on television. From a theoretical point of view, only 27.6% of papers are based on specific theoretical frameworks, while at methodological level, 77.8% rely on non-probabilistic samples and only 50% provide intercoder reliability indicators.

In terms of the theoretical framework, Bryant and Miron (2004) have pointed out, based on an analysis of 1,806 papers published by *Journalism & Mass Communication Quarterly*, *Journal of Communication* and *Journal of Broadcasting & Electronic Media* (from 1956 to 2000), that framing theory is the most prominent, followed closely behind by the agenda-setting theory. Finally, it is essential to mention the work of Lovejoy *et al.* (2016), which examined how intercoder reliability is reported in a sample of 672 papers published by *Journalism & Mass Communication Quarterly*, *Journal of Communication* and *Communication Monographs*, over a three-decade period (1985–2014). They concluded that, while Scott's *Pi* (π) is the most used intercoder reliability measure in the entire period (15%), while Krippendorff's *Alpha* (α) and Cohen's *Kappa* (k) are the most commonly used in the 2010–2014 period, with 28.8% and 25.2% respectively. They also point out that the reporting of reliability assessment ranges from 20% in 1985 to 90% in 2014, which reveals a progressive respect for this crucial step of the protocol.

The next section establishes the research objectives, questions, hypotheses, and methodological procedures that guide this research.

2. Objectives and methods

As the title of this study suggests, the purpose is to identify the ways content analysis has been used as a research method to address communication messages. This main goal is divided into two specific research objectives. The first one is to describe, by means of a bibliometric approach, the papers which use content analysis, considering such aspects as authorship and

funding. The second one is to examine specific aspects concerning the use of content analysis, such as sampling strategies and the reporting of intercoder reliability assessment, to be able to outline, in some detail, its application by the experts in the area.

To meet the proposed objectives, and in view of the findings of previous studies, the following sets of research questions and hypotheses were established. From a bibliometric point of view:

RQ1: What is the assessment of the sample of research papers in terms of average number of authors and publication language?

H1: The research presented in papers usually lacks funding.

At applied and operational level, it is interesting to contrast the following parameters:

H2a: Papers often lack a specific theoretical or conceptual framework.

H2b: Papers that do have a theoretical framework are guided by framing and agenda-setting theories.

H3: Research papers tend to resort to non-probabilistic sampling methods.

H4: Intercoder reliability reporting is the most common practice in the papers.

In terms of sampling and intercoder reliability strategies:

RQ2: Are there any differences between the journals in the sample according to their Spanish or Latin American origin?

And finally:

H5: The preferred object of study in papers using content analysis is the message.

H6: The press is the most-commonly analysed medium.

RQ3: Which topics are the most attractive to content analysts?

To answer these questions and test the hypotheses, a methodological triangulation was implemented (Denzin, 2012), articulating a bibliometric review with content analysis itself. The sampling design, one of the key elements in any empirical study (Igartua, 2006), is multi-stage stratified sampling (Neuendorf, 2017) in this case. In the first phase, we selected Spanish-language journals –from Spain and Latin America– with the highest impact factor in 2017, in the “communication” category of the *SCImago Journal & Country Rank*³. In order to be included in the sample, the journals had to be listed in the first two quartiles⁴. This criterion yielded a total of 7 publications, which in hierarchical order are: *Comunicar*, *El Profesional de la Información*, *Communication & Society*, *Revista Latina de Comunicación Social*, *Cuadernos.info*, *Comunicación y Sociedad* and *Palabra Clave*. To confer a temporal perspective on the sample, a five-year period was calculated from 2017 backwards. However, this study is not longitudinal, like most of the studies mentioned above, as it aims to carry out a cross-sectional analysis to identify, at this recent stage, the research projects that have used content analysis.

Having selected the journals, which served as data collection units, the following step was to search for the terms “*análisis de contenido*” and “content analysis” in the publications’ internal search engines. Although the search focused on the papers’ titles, abstracts, and keywords, the search engines also yielded manuscripts that included these terms in the body of the text. Therefore, and in view of the limitations inherent in this initial procedure, which did not always yield accurate results, all the papers published by the selected journals in the selected period ($N = 1,548$) were reviewed to identify precisely those that actually used content analysis. Following this sampling strategy, the total number of analysis units increased to $n = 262$ papers, representing ~16% of all those published (see: Piñeiro-Naval & Morais, 2019). This percentage is not so far, for example, from the ~25% represented by the papers detected, although over a much wider period, in the classic work by Riffe and Freitag (1997).

³ Retrieved from <http://bit.ly/36MJFoK>.

⁴ It should be noted that, according with the most recent requirements set by the Spanish National Agency for Quality Assessment and Accreditation (ANECA) for full-time university professor positions, we considered level-1 journals as those located in the first two quartiles of the *SCImago Journal & Country Rank*.

2.1. Analysis instrument and procedure

An evaluation mechanism was designed in accordance with the objectives and drawing inspiration from similar studies (Borah, 2017; Caffarel-Serra, Ortega-Mohedano & Gaitán-Moya, 2017; Escribá & Cortiñas, 2013; Fernández-Quijada & Masip, 2013; Goyanes, Rodríguez-Gómez & Rosique-Cedillo, 2018; Kim *et al.*, 2017; Martínez-Nicolás, Saperas & Carrasco-Campos, 2019; Piñeiro-Naval & Mangana, 2018; Rodríguez-Gómez, Goyanes & Rosique-Cedillo, 2018; Saperas & Carrasco-Campos, 2018; Walter, Cody & Ball-Rokeach, 2018). The instrument considered the following variables applied to each analysis unit:

Table 2: Study variables.

No./Variable	Description/Values	No./Variable	Description/Values
A. Basic identification variables			
1. Paper's publication year	From 2013 to 2017	3. Quartile in <i>SJR/Scopus</i>	0=no quartile, 1=quartile 1, 2=quartile 2, 3=quartile 3, 4=quartile 4
2. Publishing Journal	1= <i>Comunicar</i> , 2= <i>EPI</i> , 3= <i>Comm. & Soc.</i> , 4= <i>RLCS</i> , 5= <i>Cuadernos.info</i> , 6= <i>Comun. y Soc.</i> , 7= <i>Palabra Clave</i>	4. Impact factor in <i>SJR/Scopus</i>	Exact impact factor of journal containing the paper in the corresponding year
B. Bibliometric Variables		C. Analytical Variables	
5. Paper's language	1=Spanish, 2=Spanish/English, 3=English, 4=Portuguese, 5=Portuguese/English	12. Method of analysis ($\alpha_k = \text{constant}$)	1>manual content analysis, 2=automated content analysis
6. First author's affiliation	1=University ⁵ , 2=Company, 3=Public entity	13. Supporting theory/concept ($\alpha_k = 0.73$)	See Table 3
7. First author's country of affiliation	1=Spain, 2=Chile, 3=Mexico, 4=Colombia, 5=other	14. Study sample ($\alpha_k = 0.81$)	1=probabilistic, 2=non-probabilistic
8. Number of authors	Exact number of authors was recorded	15. Intercoder reliability ($\alpha_k = 0.89$)	0=not reported, 1=reported
9. Author's discipline	1=Communication, 2=Journalism, 3=Advertising & PR, 4=Political Science, 5=Sociology, 6=Psychology, 7=Education, 8=Marketing, 9=Library Science, 10=Computer Science	16. Main object of study ($\alpha_k = 1$)	1=source, 2=message, 3=audience, 4=policies/structure
10. Paper's funder	0=unfunded, 1=public entity, 2=private entity	17. Medium under analysis ($\alpha_k = 0.91$)	See Table 3
11. Funder's scope of action	0=unfunded, 1=local, 2=regional, 3=national, 4=international	18. Paper's general theme ($\alpha_k = 0.80$)	See Table 3

Source: Own elaboration.

⁵ In the case of university affiliation, only the first author's university name was recorded.

A total of 18 different variables were measured: 4 basic identification variables (one nominal and polychotomous, two ordinal and one ratio), 7 bibliometric variables (5 nominal and polychotomous, one ordinal and one ratio) and 7 analytical variables (two nominal and polychotomous, one dummy and the rest polychotomous). It is important to note, in this regard, that the values corresponding to variables 3 and 4 were extracted –as independent variables– from the *SCImago Journal & Country Rank*, which facilitated their subsequent triangulation (Denzin, 2015) with the data collected here.

Finally, it is also important to mention that the data were collected in November and December 2018 by two members of the same research group: the author of this paper, who coded the sample of 262 papers, and a collaborator trained in advance for this task. For the reliability check, a random subsample of ~10% of the cases ($n = 27$) was selected and coded by the collaborator, which made it possible to compare the two judges' data. The statistical measure used to calculate reliability was Krippendorff's *Alpha* (Krippendorff, 2011, 2017), which was estimated using the "*Kalpha macro*" (Hayes & Krippendorff, 2007) in SPSS (version 24). The average reliability of the 7 analytical variables was very satisfactory: $M(\alpha_k) = 0.85$, with values ranging from "0.73" to "1."

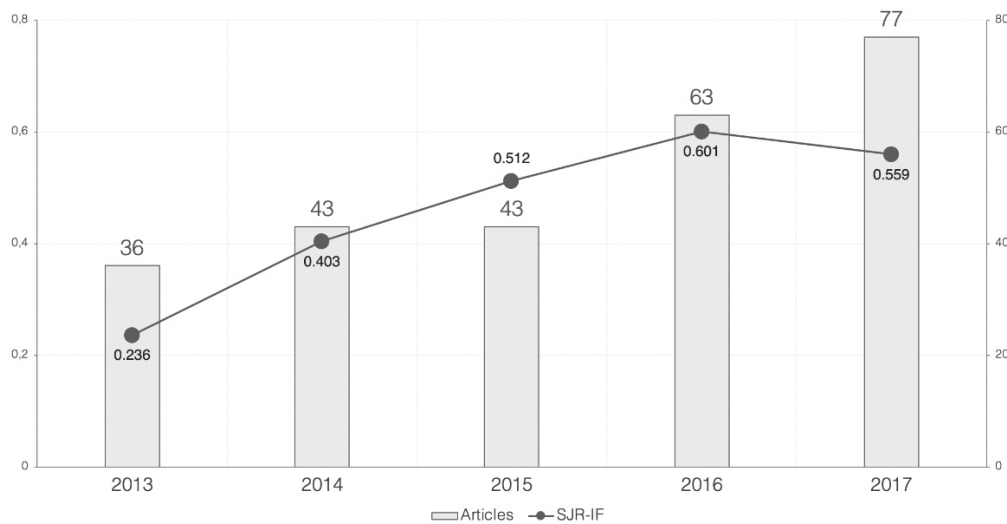
3. Results

Like the data collection instrument, the results section is divided into two blocks: first, findings corresponding to the bibliometric analysis and, second, findings relating to the use of content analysis in the papers.

3.1. Bibliometric perspective

First of all, the annual evolution of the number of papers that use content analysis is represented in Figure 1, which also shows the progression of the average impact factor:

Figure 1: Annual evolution of papers and their average impact factor.



Source: Own elaboration.

As shown by the timeline, there is a steady and pronounced increase, since the number of papers published in 2017 more than doubles the number of articles in 2013 ($M_{Annual} = 52.4$ papers, $SD = 17.05$). With regard to the *SJR* impact factor, the descriptive parameters for the sample are: $M_{SJR-IF} = 0.491$ ($SD = 0.261$). The growth experienced from the first year of analysis to 2016 is very noticeable. However, this trend is reversed in 2017 [$F_{SJR-IF \times Year}(4, 257) = 17.64$, $p < 0.001$, $\eta^2 = 0.215$]. To be precise, the main differences occur between 2016 and 2013 [$t(97) = 6.136$, $p < 0.001$, $d = 1.405$], which would be labelled as "high" depending on the size of the

effect (Cohen, 1988; Johnson *et al.*, 2008). On the other hand, correlation between the number of papers per year and their average impact factor shows a positive and statistically significant trend: $r(260) = 0.413$, $p < 0.001$; which reveals that the higher the number of published papers, the greater the impact.

The percentage distribution of the papers published by journals in the sample is as follows: *Comunicar* (8%), *El Profesional de la Información* (17.2%), *Communication & Society* (21.8%), *Revista Latina de Comunicación Social* (28.6%), *Cuadernos.info* (13%), *Comunicación y Sociedad* (5.3%) and *Palabra Clave* (6.1%). Thus, the journal with largest percentage of papers using content analysis is *Revista Latina de Comunicación Social*, followed by *Communication & Society*. Together, these two journals publish more than half of the papers in the sample (50.4%). It should also be noted that 22.5% of these papers were included in the first quartile of the *SJR/Scopus* ranking, 55.7% in the second, 16% in the third, 2.3% in the fourth and 3.4% were not included in these rankings at the time of publication. Thus, the modal quartile is the second and, in the aggregate, 78.2% of the papers exceed the average value.

The distribution of the publication languages of the papers (RQ₁) is as follows: Spanish (24.8%), Spanish/English (59.5%), English (14.1%), Portuguese (1.1%) and Portuguese/English (0.4%). Thus, 74% of the papers are written in the language of science, English, either exclusively or accompanied by a version in Spanish or Portuguese. Most authors are affiliated with universities (99.2%). The top ten most prolific universities are as follows:

1. Universidad Rey Juan Carlos ($n = 17$)
2. Universidad de Valladolid ($n = 16$)
3. Universidad del País Vasco ($n = 14$)
4. Universidad Autónoma de Barcelona ($n = 13$)
5. Universidad Complutense de Madrid and Universidad de Málaga ($n = 12$)
6. Universidad Pompeu Fabra ($n = 11$)
7. Universidad de Valencia ($n = 10$)
8. Universidad Internacional de La Rioja ($n = 8$)
9. Universidad de Salamanca and Universidad de Vigo ($n = 7$)
10. Universidad de A Coruña ($n = 6$)

Of the total of 92 universities that were identified, the 10 above –all from Spain– are the ones that produce more than half of the research papers (50.76%). In terms of countries, Spain is the one with the highest percentage of authors in the sample (74%), followed closely by Mexico and Chile, with 7.3% and 6.5% respectively.

Regarding authorship (RQ₁), the average number of authors per paper is: $M_{Authors} = 2.44$ ($SD = 1.11$). However, it should be noted that the modal value is “3” –specifically, in 37.8% of cases. With regards to disciplines, communication (42.4%), journalism (37.8%) and advertising and PR (8%) are the ones represented in 88.2% of papers, leaving a small margin for contributions from other areas, such as political science (3.4%), psychology (1.5%) and library science (1.1%). Finally, in terms of reported funding (H₁) in the sample of papers, the distribution is as follows: unfunded (56.1%), local (7.6%), regional (3.8%), national (30.2%) and international (2.3%).

3.2. Applied perspective

From a methodological point of view, 92% of papers use the manual mode of content analysis, while the remaining 8% use the automated version. In addition, Table 3 presents the distribution of papers according to 3 variables: supporting theories and concepts, media and platforms analysed, and themes covered.

Table 3: Theories, media and themes covered in the papers.

Theories/Concepts	%	Media/Platforms	%	Themes	%
Framing	18.3	Press	32.1	Politics	23.7
Agenda-setting	10.3	Social networks	22.5	General information	14.9
Web 2.0	7.6	Internet/ICT	9.9	Gender	7.3
Engagement	3.8	Television	9.5	Corporate communication	6.5
Web Design	3.4	Scientific papers	6.9	State of academia	6.5
Corporate social responsibility/Branding	3.4	Advertisements	6.1	Persuasive/advertising communication	5.3
Info/Politainment	3.1	Film	4.2	Education and sports	5
Social identity theories	3.1	Media in general	3.8	Digital society	5
Health communication	2.3	Radio	1.5	Economy	4.6
Media ecology	1.9	Internships	0.8	Culture and tourism	3.8
Transmedia	1.5	Journals	0.8	Health and sanitation	3.4
Media literacy	1.1	Cultural industries	0.4	Social movements	3.1
Selective exposure	1.1	Receptors	0.4	Environment	2.7
Cultivation theory	0.8	Repositories	0.4	Migration and minorities	2.7
Priming	0.8	Video games	0.4	Uses and audience motivations	2.7
Narrative persuasion models	0.4	Other media and platforms	0.3	War conflicts	2.3
Total	62.9	Total	100	Total	99.5

Source: Own elaboration.

The proportion of studies that explicitly refer to a specific theory (H_{2a}) or conceptual notion is 68.3%. The most common theories are two of the classical ones (H_{2b}): Framing Theory (18.3%) and Agenda-Setting Theory (10.3%). As for the object of study (H₃), the papers in the sample are highly focused on communication messages (94.3%). In this sense, the preferred media platforms (H₆) are the press (32.1%), both traditional and digital, and social networks (22.5%). At thematic level (RQ₃), politics (23.7%) is the favourite topic among media analysts. In terms of sampling methods, 34.4% of the papers employ probabilistic samples, while the remaining 65.6% opt for non-probabilistic strategies (H₃). The following contingency table relates the type of content analysis –manual vs. automated– with the sampling method and, on the other hand, the nationality of journals –Spain vs. Latin America– with the sampling method and intercoder reliability report (RQ₂):

Table 4: Cross table with multiple comparisons according to journal sample type and nationality (% column).

Content analysis methods	Total %	Sample type	
		Probabilistic	Non-probabilistic
Manual	92	84.4–	95.9+
Automated	8	15.6+	4.1–
<i>n</i>	262	90	172
Sample type	Total %	Journal nationality	
		Spain	Latin America
Probabilistic	34.4	30.3–	46.9+
Non-probabilistic	65.6	69.7+	53.1–
<i>n</i>	262	198	64
Intercoder reliability			
Reported	24.1	18.7–	40.7+
Not reported	75.9	81.3+	59.3–
<i>n</i>	241	182	59

Note: – Statistically low value (analysis of corrected typified residuals). + Statistically high value (analysis of corrected typified residuals). Source: Own elaboration.

As shown in Table 4, automated content analyses use representative samples in greater proportions [$\chi^2(1, n = 262) = 15.57, p < 0.01, v = 0.201$] than manual content analyses do. The last one of the variables in the codebook has to do with reporting of intercoder reliability. Surprisingly, 69.8% of the papers do not report any kind of statistic about intercoder agreement. Only 22.1% take into account this peculiarity of manual content analysis (H₁). On the other hand, Latin American journals publish research papers with representative samples [$\chi^2(1, n = 262) = 5.89, p < 0.05, v = 0.15$] and intercoder reliability reports to a greater extent than Spanish journals [$\chi^2(1, n = 241) = 11.79, p < 0.01, v = 0.221$].

4. Discussion and conclusions

The substantial review of the study's findings provides an overview of the uses and applications –and even practices– of content analysis in recent communication research in the Spanish-speaking world. First, there is a constant and very pronounced evolution in the period under analysis with regards to use of this technique, which attests to its growing prominence and impact. On the other hand, most authors come from the Spanish university sector, where a small group of institutions are responsible for most scientific production. Most works belong to the areas of communication and journalism and often lack funding. In the absence of financial support, it has been confirmed that content analysis is a recurring technique for the production of empirical research. To end the bibliometric analysis of the results, it should be noted that multiple authorship is the modal practice and that the internationalisation of research through publication in English –exclusively or accompanied by a Spanish version– occurs in three quarters of the sample.

From the point of view of executing the content analysis, most of the works are based on specific theoretical frameworks, particularly classical theories such as the agenda-setting theory and, above all, framing theory, which has been one of the most popular conceptual paradigms in recent decades (Piñeiro-Naval & Mangana, 2019; Saperas & Carrasco-Campos, 2015; Vicente-Mariño & López-Rabadán, 2009). Without a doubt, this circumstance represents a strength of the research carried out in the Spanish-speaking world, since linking empirical works with specific conceptual frameworks contributes, as Riffe and Freitag (1997) have pointed out, to advancement of the field. However, the fact that the samples used in these works tend to be non-probabilistic is a weakness, since their findings cannot be extrapolated to the overall population under study. However, the most problematic weakness is the lack of intercoder reliability reports, the most critical step in implementation of content analysis. According to the best practices proposed by Lacy *et al.* (2015) as well as other authors listed in the introduction, the recurrent absence of indicators –preferably Krippendorff's *Alpha*– that confirm the objectivity of the coding process is striking. As for these parameters –sampling and reliability– Latin American journals are generally more careful than Spanish ones, which is a threat that should be addressed.

Finally, the message is undoubtedly the preferred object of study. Therefore, despite the inferential nature of the method, which allows us to infer the conditions of production and reception of messages, it is information itself that arouses the greatest interest of content analysts and, more specifically, the political messages published in the press –both traditional and digital– and those posted on social networks. The following table summarises all the research questions and hypotheses of this study –classified according to their bibliometric or applied nature– their approach and their subsequent resolution:

Table 5: Summary of research questions and hypotheses divided according to their bibliometric or applied nature.

Questions/Hypothesis	Approach	Resolution
Bibliometric level		
RQ ₁	What is the assessment of the sample of research papers in terms of average number of authors and publication language?	<i>M</i> = 2.44 authors <i>M_o</i> = 3 authors English = 74%
H ₁	Unfunded research works > funded research works	Accepted
Applied level		
H _{2a}	Studies without specific theoretical frameworks > studies with specific theoretical frameworks	Rejected
H _{2b}	Most recurring theoretical framework: Framing Theory > Agenda-Setting > other theories/concepts	Accepted
H ₃	Non-probabilistic samples > probabilistic samples	Accepted
H ₄	Presence of reliability report > absence of reliability report	Rejected
RQ ₂	Are there any differences between the journals in the sample according to their Spanish or Latin American origin?	Latin American > Spanish
H ₅	Preferred object of study: message	Accepted
H ₆	Most recurrent media platform of study: the press	Accepted
RQ ₃	What is the most frequently discussed topic?	Politics

Source: Own elaboration.

As possible future lines of work, and although manual content analysis remains one of the main methods used in communication sciences, we should not overlook the fact that, in a fully digitalised media ecosystem, the volume of data that can be processed is growing in leaps and bounds, which will certainly drive automated content analysis to acquire more and more importance and reputation (Boumans & Trilling, 2016; Trilling & Jonkman, 2018). On the other hand, despite not being considered a research variable in our study, it has been noted that the authors in the sample often carry out data analyses that are restricted to the single-variable reporting of percentages or parameters of central trend. This is another aspect that could gain greater sophistication through the use of multivariate statistics, which would facilitate the checking of more complex hypotheses and, consequently, advancement of the discipline.

Finally, it is worth noting that future bibliometric and, in particular, meta-analyses are needed for the academic community to be able to propose more appropriate “public policies on research in our field” (Caffarel-Serra, 2018, p. 295), and monitor research work routines to identify new strengths, weaknesses, opportunities and threats.

This article is part of a project funded by the Portuguese Foundation for Science and Technology (Ref.: SFRH/BPD/122402/2016) and the “LabCom – Communication and Arts” R&D Unit of the University of Beira Interior (Ref.: UIDB/00661/2020).

References

- Bardin, L. (1986). *Análisis de contenido*. Madrid: Akal.
- Berelson, B. (1952). *Content analysis in communication research*. Glencoe: Free Press.
- Berger, A. A. (2016). *Media and communication research methods. An introduction to qualitative and quantitative approaches (4th Ed.)*. Thousand Oaks, CA: Sage.
- Bermejo-Berros, J. (2014). Evolución de los paradigmas, metodologías y campos de la comunicación en *Revista Latina de Comunicación Social* durante la década 2004-2013. *Revista Latina de Comunicación Social*, 69, 330-353. <https://www.doi.org/10.4185/RLCS-2014-1014>
- Borah, P. (2017). Emerging communication technology research: Theoretical and methodological variables in the last 16 years and future directions. *New Media & Society*, 19(4), 616-636. <https://www.doi.org/10.1177/1461444815621512>

- Bos, W. & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, 31(8), 659-671. [https://www.doi.org/10.1016/S0883-0355\(99\)00032-4](https://www.doi.org/10.1016/S0883-0355(99)00032-4)
- Boumans, J. W. & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, 4(1), 8-23. <https://www.doi.org/10.1080/21670811.2015.1096598>
- Bryant, J. & Miron, D. (2004). Theory and research in mass communication. *Journal of Communication*, 54(4), 662-704. <https://www.doi.org/10.1111/j.1460-2466.2004.tb02650.x>
- Caffarel-Serra, C. (2018). La metainvestigación en comunicación, una necesidad y una oportunidad. *adComunica*, 15, 293-295. <https://www.doi.org/10.6035/2174-0992.2018.15.16>
- Caffarel-Serra, C., Ortega-Mohedano, F. & Gaitán-Moya, J. A. (2017). Investigación en Comunicación en la universidad española en el periodo 2007-2014. *El Profesional de la Información*, 26(2), 218-227. <https://www.doi.org/10.3145/epi.2017.mar.08>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Earlbaum.
- Deacon, D., Pickering, M., Golding, P. & Murdock, G. (1999). *Researching communications. A practical guide to methods in media and cultural analysis*. London: Arnold Publishers.
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), 80-88. <https://www.doi.org/10.1177/1558689812437186>
- Denzin, N. K. (2015). Triangulation. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (pp. 1-6). NJ: Wiley & Sons. <https://www.doi.org/10.1002/9781405165518.wbeosto50.pub2>
- Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International*, 13(3), 313-321. <https://www.doi.org/10.1080/07399339209516006>
- Escribà, E. & Cortiñas, S. (2013). La internacionalización y las coautorías en las principales revistas científicas de Comunicación en España. *Comunicar*, 21(41), 35-44. <https://www.doi.org/10.3916/C41-2013-03>.
- Espín, J. V. (2009). El análisis de contenido: una técnica para explorar y sistematizar información. *Revista de Educación*, 4, 95-105.
- Fernández-Quijada, D. & Masip, P. (2013). Tres décadas de investigación española en comunicación: hacia la mayoría de edad. *Comunicar*, 21(41), 15-24. <https://www.doi.org/10.3916/C41-2013-01>
- Goyanes, M., Rodríguez-Gómez, E. F. & Rosique-Cedillo, G. (2018). Investigación en comunicación en revistas científicas en España (2005-2015): de disquisiciones teóricas a investigación basada en evidencias. *El Profesional de la Información*, 27(6), 1281-1291. <https://www.doi.org/10.3145/epi.2018.nov.11>
- Johnson, B. T., Scott-Sheldon, L. A. J., Snyder, L. B., Noar, S. M. & Huedo-Medina, T. B. (2008). Contemporary approaches to meta-analysis in communication research. In A. F. Hayes, M. D. Slater & L. B. Snyder (Eds.), *The SAGE Sourcebook of Advanced Data Analysis Methods for Communication Research* (pp. 311-347). Thousand Oaks, CA: Sage.
- Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. <https://www.doi.org/10.1080/19312450709336664>
- Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage. <https://www.doi.org/10.4135/9781412985451>
- Holsti, O. R. (1968). Content Analysis. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (pp. 596-692). Oxford: Addison-Wesley.
- Igartua, J. J. (2006). *Métodos cuantitativos de investigación en comunicación*. Barcelona: Bosch.
- Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLOS Biology*, 16(3), e2005468. <https://www.doi.org/10.1371/journal.pbio.2005468>
- Kim, Y., Kim, B., Kim, Y. & Wang, Y. (2017). Mobile communication research in communication journals from 1999 to 2014. *New Media & Society*, 19(10), 1668-1691. <https://www.doi.org/10.1177/1461444817718162>

- Krippendorff, K. (1990). *Metodología de análisis de contenido. Teoría y práctica*. Barcelona: Paidós Comunicación.
- Krippendorff, K. (2004). Reliability in content analysis. Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433.
<https://www.doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures*, 5(2), 93-112. <https://www.doi.org/10.1080/19312458.2011.568376>
- Krippendorff, K. (2013). Commentary: A Dissenting View on So-Called Paradoxes of Reliability Coefficients. *Annals of the International Communication Association*, 36(1), 481-499.
<https://www.doi.org/10.1080/23808985.2013.11679143>
- Krippendorff, K. (2017). Reliability. In J. Matthes, C. S. Davis & R. F. Potter (Eds.), *The International Encyclopaedia of Communication Research Methods* (pp. 1-28). NJ: Wiley & Sons. <https://www.doi.org/10.1002/9781118901731.iecrm0210>
- Lacy, S., Watson, B. R., Riffe, D. & Lovejoy, J. (2015). Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly*, 92(4), 791-811.
<https://www.doi.org/10.1177/1077699015607338>
- Lombard, M., Snyder-Duch, J. & Campanella, C. (2002). Content analysis in mass communication. Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587-604. <https://www.doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Lovejoy, J., Watson, B. R., Lacy, S. & Riffe, D. (2016). Three decades of reliability in communication content analyses: Reporting of reliability statistics and coefficient levels in three top journals. *Journalism & Mass Communication Quarterly*, 93(4), 1135-1159.
<https://www.doi.org/10.1177/1077699016644558>
- Lozano, J. C. (1994). Hacia la reconsideración del análisis de contenido en la investigación de los mensajes comunicacionales. In C. Cervantes & E. Sánchez (Eds.), *Investigar la comunicación: propuestas iberoamericanas* (pp. 135-157). México: Universidad de Guadalajara, CEIC, ALAIC.
- Martín-López, E. (1963). El análisis de contenido. *Revista de estudios políticos*, 132, 45-64.
- Martínez-Nicolás, M., Saperas, E. & Carrasco-Campos, Á. (2019). La investigación sobre comunicación en España en los últimos 25 años (1990-2014). Objetos de estudio y métodos aplicados en los trabajos publicados en revistas españolas especializadas. *Empiria. Revista de metodología de ciencias sociales*, 42, 37-69.
<https://www.doi.org/10.5944/empiria.42.2019.23250>
- Naccarato, J. L. & Neuendorf, K. A. (1998). Content Analysis as a Predictive Methodology: Recall, Readership, and Evaluations of Business-to-Business Print Advertising. *Journal of Advertising Research*, 38, 19-29.
- Neuendorf, K. A. (2009). Reliability for content analysis. In A. B. Jordan, D. Kunkel, J. Manganello & M. Fishbein (Eds.), *Media messages and public health: A decisions approach to content analysis* (pp. 67-87). New York: Routledge.
- Neuendorf, K. A. (2017). *The content analysis guidebook (2nd Ed.)*. Thousand Oaks, CA: Sage.
- Piñeiro-Naval, V., Igartua, J. J., Marañón-Lazcano, F. J. & Sánchez-Nuevo, L. A. (2018). El análisis de contenido y su aplicación a entornos web: un caso empírico. In C. Caffarel, J. A. Gaitán, C. Lozano & J. L. Piñuel (Eds.), *Tendencias metodológicas en la investigación académica sobre Comunicación* (pp. 253-272). Salamanca: Comunicación Social.
- Piñeiro-Naval, V. & Mangana, R. (2018). Teoría del encuadre: Panorámica conceptual y estado del arte en el contexto hispano. *Estudios sobre el Mensaje Periodístico*, 24(2), 1541-1557.
<https://www.doi.org/10.5209/ESMP.62233>
- Piñeiro-Naval, V. & Mangana, R. (2019). La presencia del *framing* en los artículos publicados en revistas hispanoamericanas de comunicación indexadas en Scopus. *Palabra Clave*, 22(1), e2216. <https://www.doi.org/10.5294/pacla.2019.22.1.6>

- Piñeiro-Naval, V. & Morais, R. (2019). Study of the academic production on communication in Spain and Latin America. *Comunicar*, 27(61), 108-117. <https://www.doi.org/10.3916/C61-2019-10>
- Piñuel, J. L. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Estudios de Sociolingüística*, 3(1), 1-42.
- Riffe, D. & Freitag, A. (1997). A content analysis of content analyses: Twenty-five years of Journalism Quarterly. *Journalism & Mass Communication Quarterly*, 74(3), 515-524. <https://www.doi.org/10.1177/107769909707400306>
- Riffe, D., Lacy, S. & Fico, F. (2014). *Analyzing Media Messages. Using Quantitative Content Analysis in Research (3rd Ed.)*. New York: Routledge.
- Rodríguez-Gómez, E. F., Goyanes, M. & Rosique-Cedillo, G. (2018). Communication Research in Spain: Labor Temporality, Intensive Production and Competitiveness. *Communication & Society*, 31(4), 229-242. <https://www.doi.org/10.15581/003.31.4.229-242>
- Saperas, E. & Carrasco-Campos, Á. (2015). The operationalization of the concept of “framing” in the “Journal of Communication” (2009-2013): objects of study, research techniques and theoretical construction. *Communication & Society*, 28(4), 49-66. <https://www.doi.org/10.15581/003.28.4.sp.49-66>
- Saperas, E. & Carrasco-Campos, Á. (2018). Journalism research: a dominant field of communication research in Spain. A meta-research on Spanish peer-reviewed journals (2000-2014). *Estudos em Comunicação*, 26(1), 281-300. <https://www.doi.org/10.20287/ec.n26.v1.a16>
- Saperas, E. & Carrasco-Campos, Á. (2019). ¿Cómo investigamos la comunicación? La meta-investigación como método para el estudio de las prácticas de investigación en los artículos publicados en revistas científicas. In F. Sierra Caballero & J. Alberich Pascual (Eds.), *Epistemología de la comunicación y cultura digital: retos emergentes* (pp. 217-230). Granada: Universidad de Granada.
- Trilling, D. & Jonkman, J. G. (2018). Scaling up content analysis. *Communication Methods and Measures*, 12(2-3), 158-174. <https://www.doi.org/10.1080/19312458.2018.1447655>
- Vicente-Mariño, M. & López-Rabadán, P. (2009). Resultados actuales de la investigación sobre *framing*: sólido avance internacional y arranque de la especialidad en España. *Zer. Revista de Estudios de Comunicación*, 14(26), 13-34.
- Walter, N., Cody, M. J. & Ball-Rokeach, S. J. (2018). The Ebb and Flow of Communication Research: Seven Decades of Publication Trends and Research Priorities. *Journal of Communication*, 68(2), 424-440. <https://www.doi.org/10.1093/joc/jqx015>
- White, M. D. & Marsh, E. E. (2006). Content analysis: a flexible methodology. *Library Trends*, 55(1), 22-45.
- Wimmer, R. D. & Dominick, J. R. (2011). *Mass Media Research: An Introduction (9th Ed.)*. Boston: Wadsworth.
- Zamith, R. & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 307-318. <https://www.doi.org/10.1177/0002716215570576>
- Zhao, X., Liu, J. S. & Deng, K. (2013). Assumptions behind Intercoder Reliability Indices. *Annals of the International Communication Association*, 36(1), 419-480. <https://www.doi.org/10.1080/23808985.2013.11679142>