Street Art & Urban Creativity | Vol. 11, No. 4, 2025 | ISSN 2183-9956
Street Art & Urban Creativity. Scientific Journal
https://doi.org/10.62161/sauc.v11.5729

CONTEMPORARY ART IN EL PAÍS NEWSPAPER: A DUAL ANALYSIS OF CONTENT

COMPARISON OF RESULTS USING THE STEMPEL METHOD AND COMPUTATIONAL ANALYSIS WITH PYTHON PROGRAMMING

RUBÉN FERNÁNDEZ-COSTA O'DOGHERTY [1], ELVIRA CALVO GUTIÉRREZ [1], JOAQUÍN SOTELO GONZÁLEZ[1]
[1]Universidad Complutense de Madrid, Spain

KEYWORDS

Contemporary Art
Newspaper
El País
Stempel
Quantitative
Automated Studies
Python
Web Scraping
Data Science

ABSTRACT

This study confirms the consistency of a dual repeated analysis of contemporary art content in El País newspaper, employing, first, the Stempel methodology and, second, its automation through web scraping techniques using Python programming. Pending the development of fully autonomous AI systems that will enable real-time monitoring of information system content, both approaches confirm that—marking a century since the traditionally accepted birth of contemporary art (Duchamp, 2017)—there is limited representation of current art in an iconic Spanish media outlet for a statistically typical year, compared to coverage in the Sports section and other cultural categories such as Film and Series. Additionally, the study reveals a hyperbolic and materialistic vocabulary in the representation of contemporary art.

RECEIVED: 27 / 02 / 2025
ACCEPTED: 10 / 06 / 2025

1. Introduction

The application of quantitative techniques to artistic and communicative fields has been gradually incorporated throughout the second half of the 20th century, until its habitual use in the Social Sciences and Humanities was consolidated (Bardin, 1986; Berelson, 1952; Holsti, 1969; Krippendorf, 1990). In honour of the recently deceased Professor Dr. Jorge Lozano, who was a member of the Faculty of Information Sciences at the Complutense University of Madrid, it can be posited that we are currently experiencing a "new Third Stage". This assertion is supported by the development of "large-scale cultural analytics" (Lozano and Martín, 2018), a methodology that facilitates the quantitative analysis of substantial data sets, thereby yielding insights that facilitate reflection (DiMaggio et al., 2013; Kristensen y From, 2015; Tilles, 2016).

The present article focuses on two considerations prior to quantitative analyses of information systems in this third stage. Firstly, there is a necessity for profound and specialised knowledge of the subject matter – in this case, the intricate domain of contemporary art and its representation in a shifting communication landscape – and a discernible ethical commitment: cultural hierarchies and historical narratives are “multi-casua”' and there are manifold implications of language in the idiosyncrasies of a sector (Bauman, 2007; Bauman, 2018; Purhonen et al., 2019). The present study does not refute this fact but rather serves to emphasise it. Secondly, it is essential to confirm not only the feasibility of the Software Studies in terms of data accessibility, but also the quality and robustness of the metrics, making use of statistics to support the hypotheses with a subsequent qualitative review.

It is widely accepted that the media, specifically newspapers, contribute to shaping public opinion through the selection of topics (agenda setting), news angles and frames (news values/framing). The media are referred to as "institutions with which people think" (Jensen, 2014) because they intervene in the exchanges between different social systems (Luhmann, 2005; Luhmann, 2000). In Spain, the influence of online newspapers is particularly salient. According to the Annual Online Media Study (IAB, 2017), "within the media, those with the highest penetration are online newspapers, followed by thematic portals". The study also found that online newspapers have a "higher level of trust and credibility" compared to radio, television and thematic portals.

In the contemporary era, the concept of the “daily paper” is no longer applicable to the traditional form of the “newspaper”, which is now regarded as “continuous” and “accessible”. This transformation has occurred due to the advent of mobile devices and social networks, which have facilitated the dissemination of information (Dans, 2012, p. 57). Consequently, “the online newspaper stands out as the digital medium that Spaniards, especially [...] individuals aged 30 and over, consult most often during the week” (IAB, 2017). In Spain, elpais.com is, according to Comscore, the most widely read online media outlet (2,834,000 unique visitors), as well as a highly influential and traditional media outlet due to its relationship with culture. Similarly, during the 2019 COVD pandemic, consumption of professional news media did not decline, with Comscore reporting a 34% growth in the first quarter of 2020. This trend has since continued.

In the domain of contemporary art, 2017 signified the centenary of what is recognised, within the context of Western history, as the inaugural contemporary work (Marcel Duchamp's readymade “The Urinal” by Marcel Duchamp, created in 1917), which, in a symbolic sense, marked the commencement of a novel epoch distinguished by significant communicative and artistic liberty, situating the artist at the vanguard of determining the parameters of artistry, with an emphasis on the creator rather than on the intricacies of execution or the subject matter of the work. This Copernican shift in stance, as posited by Anna María Guasch, precipitated a complete disconnection between readers of the mass media and the most radically contemporary creations. Guasch (2007, p. 17) asserts that “the public does not recognise the art of its time as its own, an attitude that culminates in rejection or, at the very least, a certain accusation of unintelligibility, discursive opacity and dispersion". In summary, the year 2017 is a pivotal one in the context of this study due to the following reasons: a) It corresponds to the precise moment when precisely 100 years have elapsed since the birth of contemporary art (1917); b) The year under

consideration is considered to be a statistically typical year because it is preceded by a period of hyper-events, with a sharp decline in exhibitions, such as the COVD-19 pandemic (the years 2019-2022 are considered by experts to be years of "recovery" in terms of visitor numbers); c) It does not correspond to a significant anniversary or national celebration of contemporary art (although the "80th anniversary" of Picasso's birth will be acknowledged).
d) In this study, the results are confirmed by the outcomes of previous and subsequent years, which are also regarded as "typical".

This work aligns with the research published in the thesis *From Scandal to Mourning: an Analysis of News Trends on Contemporary Art in the Online Press* (Fernández-Costa, 2020), and proposes a twofold exercise: The objective of this study is to ascertain the continued validity of the communicative proposal of contemporaneity in a leading medium. This will be achieved by means of two analyses: a manual analysis and an automatic consecutive one. These analyses will be based on the following starting hypotheses: H1) low percentage of contemporary art content, compared to others such as sport; H2) maximum importance given to cinema and series in the culture section and H3) use and abuse of certain words in the way of covering current artistic creation. Furthermore, the study seeks to identify biases associated with grief, characterised by a sense of disaffection with the natural world, and the communicative formulas employed by media outlets in their reporting on this matter.

El País newspaper (PRISA Group), first published in 1976, is not only the most widely distributed general daily newspaper in Spain at the time of analysis (a category that varies according to the month), but also a medium with a long cultural tradition. In the initial phase of the study, a substantial volume of data was meticulously compiled from the *El País* newspaper archive, employing the Stempel method and the concepts of “constructed weeks” and “annual cyclicality”. As Javier Guallar (2012) states in his thesis, *Digital Press Archives. Analysis of Spanish Newspapers*, it is imperative for social science researchers that the media do not impede access to their newspaper archives.

In a subsequent section, an automatic analysis of the aforementioned issues is conducted, employing web scraping techniques based on the tagging of articles by the media. These two preliminary steps are of paramount importance in order to gain a proper understanding of the processing and monitoring of the media. They can be considered as a preparatory step prior to the arrival of fully autonomous Artificial Intelligence (AI) programmes, which are even referred to as “black box” programmes (see Harvard Consilience). These programmes can be useful in identifying biases and hallucinations in the analysis, and in shedding light on the results obtained.

2. Manual method (Stempel) of Quantitative Analysis to Monitor the Contents of Contemporary Art in the Online Edition of *El País* Newspaper

Quantitative analysis of a periodical media is a common practice within the scientific field, with the "constructed week" method being the most widely employed approach (Krippendorff, 1990; Stempel, Westley, 1989). The construction of a week is predicated on the principle of randomisation, whereby a random Monday is succeeded by a random Tuesday and so forth, until a random Sunday is reached, thus completing the sequence of seven days selected from all possible dates. The seminal study by Stempel (1952) was the first to establish that two “constructed” weeks were sufficient to represent the behaviour of a media over a full year. Subsequent studies by Davis and Turner (1951), Jones & Carter (1959), and Riffe et al. (1993), however, demonstrated that two “constructed” weeks are sufficient. Furthermore, Hester and Dougall (2007) asserted that at least doubling the data is necessary to adequately characterise an online medium.

The present study employs a rigorous analytical approach by examining the informative texts published on the newspaper's website, hereinafter referred to as "posts" or "news items". *El País* newspaper archive, as confirmed by the newspaper itself, has been digitised and made available online since 7 February 2012. This digital content is therefore now available for access alongside the print edition. The initial one-year timeframe of this analysis coincides with the typical cultural

and artistic event cycle, thereby facilitating the identification of problematic practices and positive developments.

A total of 1,271 posts were analysed manually, considering the three daily versions of the newspaper's home page (morning, afternoon and evening). It was determined that, in the absence of the Stempel method, the volume of data would be unfeasible, since 100 new posts per day would correspond to 36,500 new posts per year, a number that would be tripled by the existence of three daily editions. In this particular instance, the utilisation of stratified sampling, in conjunction with the Excel functions Randbetween and Date, has facilitated the generation of the requisite lists of dates for the manual sampling process. This process has been executed daily, encompassing each week of the year and each semester (with two dates allocated for each day of the week, including Monday, Tuesday, Wednesday, and so forth) during the year 2017, thereby encompassing the entire period from January to December. This results in a total of 1,271 posts, excluding the non-artistic posts from the afternoon and evening editions. Of these, 402 pertain to culture and 43 to contemporary art.

The classification of topics of *El País* posts is determined by the Cultural Domains categories established by UNESCO in 2009. In instances where ambiguity persists, the decision is made in opposition to the hypothesis, thereby ensuring the statistical behaviour is maintained. The following cultural categories are thus defined: "contemporary and current art", "classical, heritage or functional art (architecture and design)", "performing arts", "cinema and series", "current literature", "classical literature", "current music" and "classical music". Methodologically, the researcher has counted the posts on three occasions in each measure, excluding advertisements and banners, and the “most read” sections.

The analysis and coding sheet was developed in-house and was designed based on the work of Galtung and Ruge (1965) and Fernández del Moral (2007). The sheet consists of a total of 20 fields (some dependent on others), with variables and values of different natures, including: The following variables were analysed: “Day of the week (M-F)”, “Semester (1 or 2)”, “Week constructed (1-5)”, “Edition (morning, afternoon or evening)”, “Total number of news items in Home”, “Total number of news items on Current Art”, “News”, “Headline”, “Focus”, “Full text”, “Number of news items Cinema and Series, with their headlines”, “Number of news items Theatre and Performing Arts (including opera), with their headlines”, “Number of news items Classical Music”, The following headlines are to be considered: “Number of news items Non-Classical Music”, “Number of news items Classical Literature”, “Number of news items Current Literature”, “Number of news items Culture Section (automatic, according to their web classification)”, “Number of other news items related to non-contemporary art (classical art, heritage, architecture, etc.)”. The following data has been collated from the headlines: “Total number of cultural news items analysed each day (sum of all categories)” and “Number of news items in the Sports section”.

2.1. Results of Manual Analysis [Stempel]

The results presented in this section correspond to an observation window of one full year, a natural study period due to the cyclical nature of most news and events, which follow an annual periodicity. On average, the total number of new articles published online daily by *El País* newspaper is 90.79 items (it has been verified that other online media also publish an average close to 100 posts). The median and standard deviation confirm that the results follow a statistically typical distribution, forming a symmetrical Gaussian bell curve, with a median of 88 and a standard deviation of 8.37 with respect to the measurement of new posts per day. The counting process is conducted manually, and the human error introduced in the standard deviation of the counting process is considered statistically negligible, as successive measurements involve random errors that are assumed to balance out through overestimation and/or underestimation.

Regarding the results for the category defined in the bespoke classification as “Culture”, an average of 28.71 posts are published daily, with a median of 29 and a standard deviation of 6.18.

For news related to Contemporary Art in *El País*, the average is 3.07 articles, with a median of 3 and a standard deviation of 1.54. For the remaining cultural categories, the following results were obtained: 10.29 items in "Film and Series" (mean 8.5, median 5.85) and 6.07 in 'Literature' (mean 6.5, median 5.85). For the differential category relevant to the hypothesis, also measured for this study, namely "Sport", an average of 5.43 daily content items was recorded (mean 5, median 0.65).

As approximately 100 new content items are published daily, the percentages obtained are similar to the totals, as will be observed below. In proportional terms, the percentages for "Culture" are 31.70% (median 32.48%, standard deviation 6.94%), for "Contemporary Art" 3.40% (median 3.22%, standard deviation 1.76%), for "Film and Series" 11.35% (median 9.09%, standard deviation 6.69%), and for "Literature" 6.72% (median 7.02%, standard deviation 2.56%). For the differential analysis category of "Sport", the results are 16.40% (median 15.80%, standard deviation 0.94%).

2.2. Conclusions from the Results of the Manual Analysis [Stempel]

As previously discussed, relying on the Central Limit Theorem and given a sufficiently large sample size, we can assume that the proportions or means follow a normal or Gaussian distribution, enabling the calculation of confidence intervals to test the hypotheses.

Firstly, regarding the hypothesis (H1) of a low percentage of content referring to contemporary art and current artistic creation on the homepage of *El País*, compared to other sections such as Sport, the results show that the mean number of contemporary art news items is 3.07, with a median of 3 and a standard deviation of 1.54. On a Gaussian curve, applying the Central Limit Theorem with a 90% confidence interval, [$P \pm 1.65 * \sqrt{((p(1-p)/98))}$], the lower bound is —4.59%— which can be adjusted to zero —and the upper bound is 11.40%. Thus, the calculated mean proportion ("p") ranges at 90% confidence between 0 and 0.114, indicating that the average proportion of news related to current artistic creation is less than 11%, consistently lower than the percentage averages of other topics, including "Sport", which ranges between 15.46% and 17.34%. Consequently, the first hypothesis is confirmed.

Secondly, concerning the hypothesis (H2) that within the culture section, maximum importance is given to cinema and series, the data indicate that, in absolute terms, the category with the highest average content in *El País* is "Cinema/Series" (followed by "Modern Literature"). However, the distribution of proportions for cinema news falls between the categories of art and sport, which does not allow us to conclude that this is consistently the case on any given day, although it appears to be highly habitual.

Thirdly, regarding the hypothesis (H3) of the use and abuse of certain words in the coverage of current artistic creation, two preliminary steps were undertaken: for all collected texts, "stop words" were removed, and a comparison was made with the typical word frequency occurrences in the Spanish language according to the Corpus de Referencia del Español Actual (CREA, accessible online via the Real Academia Española, RAE). For *El País*, the analysis covered 33,905 words, with the relational analysis revealing the following characteristics:

- Hyperbolic focus: Repeated use of the word "more" and hyperbolic expressions. While in the CREA corpus of the RAE the word "more" ranks 23rd, here it tops the ranking. The aim is to capture attention using adverb-based linguistic structures such as "the most". The words "life" and "world" are the first two abstract terms in this manual analysis, and the word "time" (in the sense of "once") also indicates an iterative approach.
- Institutional narrative: Use of the words "exhibition", "museums", "room", and "work". In this case, the medium opts to include content that fits within a "hanger", i.e., aligned with an "agenda" that encourages visitors to art centres to engage in a leisure activity, often consumed as another product of the culture industry. As Groys writes, "museums have become sites for temporary exhibitions rather than spaces for permanent collections" (Groys, 2014, p. 88).

- Centralist approach: Overall, the sample provides evidence of a centralist focus (Madrid-Barcelona) and predominantly features male artists. “Picasso” appears as the first proper name on the list, influenced by the 80th anniversary of Guernica.

3. Automation of Quantitative Analysis Using Web Scraping: Contemporary Art Content in the Online Edition of *El País* Newspaper

While a manual or artisanal study relies on direct observation by the researcher, much larger quantitative results can be obtained using web scraping techniques and Python programming, in the category of what is currently known as data science or software studies.

The programming language employed in this research was Python 3.7.4, which is an interpreted, multi-paradigm, object-oriented, open-source, free programming language with libraries that is characterised by ease of debugging and flexibility (thus creating an easily debuggeable environment). In this particular instance, the Pandas and BeautifulSoup libraries were utilised, in conjunction with the general-purpose functions “urllib” and “request”. These functions can be employed in unison to facilitate web scraping and crawling. Moreover, the integrated development environment (IDE) Pycharm was employed to develop and test the code prior to execution. Additional support was provided by Atom and Sublime. Information pertaining to the technical procedure and the analysis code generated is omitted, as it is reusable at this time and is available to the academic community upon direct request.

The selection of tags that the designed programme must track is of paramount importance. For the purposes of this analysis, the most suitable tag is "Contemporary art" from the following list: The following terms are to be considered in this study: ART, Artà, Artana, Art Barcelona, Art Basel, Art Blakey, Art, Abstract Art, Arteaga, Ancient Art, arteBa, Baroque Art, Byzantine Art, Conceptual Art, Contemporary Art, Degenerate Art, Digital Art, Egyptian Art. The following categories of art are covered: artefacts created in the home, flamenco art, gothic art, Greek art, Iberian art, Islamic art, Arteixo, Arteixo Telecom, medieval art, Mesopotamian art, Artemio, Artemio Baigorri, Artemio Cuenca, Artemio Precioso, Artemi Rallo. The following areas of enquiry are to be considered: Artemi Rallo Lombarte, Artemisia Gentileschi, Mudejar art, Multimedia art, Artenara, Artenbrut, Neoclassical art, Pop art, Arte povera, Prehistoric art, Pre-Romanesque art, Religious art, Renaissance art, Romanesque art. Roman art, Artés, Crafts, Artisan shirt makers, Decorative arts, Performing arts, Graphic arts, 20th-Century art, Illegal arts, Martial arts, Artespaña, Plastic arts, TV art, Urban art, Artez, Art Futura.

As is generally accepted, the medium under scrutiny possesses a series of characteristics that determine the manner in which the analysis is conducted and are contingent on the consistency of the construction of the whole. The data obtained in CSV format is transformed into Excel tables, and graphs are produced using the R programme and the RStudio development environment. As outlined in the following section, the decision was taken to undertake a comprehensive analysis of the digital front pages of *El País* newspaper over the course of a year. The reliability of the results obtained in previous and subsequent years is then confirmed.

3.1. Results of Web Scraping Method

Using web scraping, a table representing the entire year is automatically generated. Given that El País newspaper publishes three daily editions—morning, afternoon, and evening (with web addresses ending in /m, /t, and /n, respectively)—the initial list produced by the programme contains 365 × 3 lines, equating to 1,095 “dates” for a year.

Upon execution, the programme generates 7,793 failed links out of a total of 108,547 (representing only a 7.7% loss of homepage links), primarily due to non-existent links, links without text or containing video content, or links resulting in “Error 403”. For the 108,547 posts successfully retrieved, the developed web scraping programme identifies 9,238 posts tagged with the word “culture” (8.5%) and 6,690 tagged with “sports” (6.2%).

It is confirmed that the posts appearing daily on each successive version of the “Homepage” (three per day) are, by the grid’s configuration, prominently featured each day, meaning the

"Homepage" metric characterises the medium's behaviour. Likewise, it is confirmed that web scraping captures all desired data: a total of 108,547 posts analysed, corresponding to 6,690 for "Sports", 9,238 for "Culture", 2,649 for "Cinema", 9,238 for "Art", and 143 for "Contemporary Art". In percentage terms, the categories are distributed as follows: 6.2% for "Sports", 8.5% for "Culture", 2.4% for "Cinema and Series", 1.9% for "Art", and less than 1% for "Contemporary Art".

3.2. Conclusions from Automatic Analysis: Web Scraping

In relation to the research hypotheses, the programme identified 2,064 posts containing the tag "art", i.e. approximately 2 per cent, and 143 with the tag "contemporary art", which represents a percentage of 0.13%. Even in the event of the most unfavourable of scenarios, incorporating posts with alternative tags such as "20th century art" (110) and "urban art" (62) —the two most significant tags according to the El País tag dictionary— the percentage would be less than 1% (0.29%) of those highlighted on the homepage. This would serve to demonstrate that the hypotheses are fulfilled.

In a final analysis, even if the decision is taken to include the remaining posts tagged in the category (90+172), the percentage remains below 1%. This suggests that the media outlet's coverage of "contemporary art" is not a priority objective, and is substantially lower than the other categories defined (7.02% for "sport" compared to 0.24% for "contemporary art").

This percentage includes links that are still valid at the time of sampling to El País Semanal, The Huffington Post and other publications embedded in this mosaic culture (Moles). In this case, the boxplots and histograms confirm all the measurements for the times of day, i.e. 3, morning, afternoon and evening, in a temporal succession of 1095 annual samples, which may contain repeated posts (again, a decision against the search), but the hypotheses are fulfilled in all cases.

In the following step, a comprehensive analysis of the trends identified in the aggregate volume of data collected for the newspaper El País during all editions of 2017 is to be conducted utilising the RStudio programme. For the sake of simplicity, the analysis graphs in this section are not included, but are available to the research community. The following conclusions are verified:

The number of daily posts on the home page is approximately 100, which corresponds approximately to the new content generated daily.

Posts tagged as "contemporary art" are seldom featured on the home page (the digital front page) of El País. In any case, for El País, the number of posts with the tag "culture" (left), in the broad sense, that are highlighted is of a similar order of magnitude to that of "sport" (right).

Regarding the initial two hypotheses, the proportion of "culture" posts exhibits a variation in accordance with the date, in accordance with the anticipated trends. For instance, the maximum number of articles categorised as "culture" on 26 November corresponds to the special coverage of the Guadalajara Book Fair, in which El País participated. Furthermore, during the months of November and December, there was an increase in the presence of general cultural content on the front page. A seasonal pattern is evident in the category of "Contemporary art", coinciding with the ARCO art fair. The first six peaks in the table are in February, with the start of the season also marked in September. However, a comparison between Culture and Contemporary art confirms that the latter accounts for a much smaller percentage than other cultural offerings.

In order to resolve the third hypothesis, an analysis was conducted of a consecutive list of post headlines related to contemporary art during 2017, with a view to describing the "construction of discourse" with "biases" or "hooks". The predominant bias identified is "reduction to absurdity", a concept analogous to the notion of "scandal", which is evident in nearly half of the headlines.

Table 1. Headlines found by tag "arte contemporáneo" (contemporary art) in *El País* 2017

Between Lenin without a body and Franco without a head
The mysterious beheading scene that Brussels is talking about
Manual of acupuncture
Resetting' Abstract Expressionism
That road that leads to a star

Argentina shakes off nostalgia at Arco 2017
David Lynch's haunting strangeness
In praise of cut and paste
The dizzying institutional critique
In the same boat
The moving wall
When Warhol visited Spain for the last time
Arco boosts optimism
The King and Queen of Spain and the Argentine President inaugurate Arco 2017
Could even a child create contemporary art? The little ones at Arco 2017 respond
Isa Calderón and Amarna Miller take a porn quiz at ARCO
Julian Rosefeldt: "We live in an age of cultural masturbation".
Seville opens CaixaForum
That was all
When the rumour materialises and Obama turns up in a museum
The banlieue relies on art to counter marginalisation
Poets in the museum
Under the cobblestones, the museum
Martí Fluxá replaces Guillermo de la Dehesa on the Reina Sofía Board of Trustees
Versailles in the down-town
Luis Barragán's ashes arrive in Mexico turned into a diamond
Miquel Barceló installs his Noah's Ark in Salamanca
Another way of being politicians
Venice Biennale: let the art games begin!
The Biennale En Marche (LBM)
Basquiat beats auction record for an American artist with 99 million euros
Art on paper
As in “Sherlock Holmes”, they were all lying.
A daughter of the great masturbator?
A treasure trove of art heads for Spain
Arroyo ascends to the Olympus of art on the Côte d'Azur
Sonia and the absolute future
The great art centre emerging from Valencia's industrial decay
Bleda y Rosa: "We look towards smaller, fragile geographies".
How much do “Las Meninas” cost?
He preferred not to do so
Richard Serra: "The best art is intrinsically useless".
Screens on the run
Painting with concrete feet
When the muse is an emoji
Truth as an essay
San Sebastian: art, food, beach and controversy
Instagrameables or the assault of images
Doris Salcedo: "The difficult thing is to achieve an invisible image, a subtle iconography".
When sex was too explicit
And Masotta committed a “happening”.
Kusama, the Japanese queen of polka dots
"Painting in Arabic is not hostile, it is poetry, and that calls for dialogue".
Morocco is impregnated with Spanish art
José Luis Alexanco: "Art has become a spectacle".
Naked bodies and feminism, this is the universe of Maria Hesse.
Belgian artist freed after 19 days chained to marble slab
Eduardo Arroyo, 21st century
Yoko Ono installs a ladder to climb to the sky of Cordoba

Source: Own elaboration with *webscrapping* in *Python,* 2024

For the entire corpus of 76 complete headlines (57,061 words) pertaining to contemporary art that appeared on the front pages of the newspaper throughout the year, a compression operation was conducted (involving the elimination of stop words) and the absolute and relative frequencies of occurrence were calculated. The following conclusions were reached:

- The "conservative" approach involves the initial technique of painting with a FAPT (absolute frequency of occurrence of a word in a text or set) of 39 (painter, 18), followed by sculpture (21), installation (19) and photography (17). The concept of exhibition or display (53+78) is accorded the utmost significance, eclipsing the permanent collection (39) and aligning closely with the themes of "agenda" and "free time". The focus of the discourse is predominantly oriented towards the events and activities that transpire within the confines of the museum (112) or museums (46), as opposed to the spatial dimensions of these institutions (44) or the plural form of spaces (15). A pronounced centralist bias is evident, with the Reina (23) emerging as the favoured locale and Madrid (51) significantly outranking Barcelona (16).
- The linguistic construction under scrutiny here seeks what has been termed "hyperbolic intensification", with comparative phraseology using the adverb "most" (292), but also "new" (71), "much" (49) and "even" (108); also "never" (28) and "always" (49). With regard to the scandal argument, despite the discrepancy in the lexicon in this area, the use of the terms sexual (11), sex (8) and sexuality (9) is particularly salient. With regard to the commercial argument, the market has a fact of 23 million, which is almost double that of 41 million, in addition to 32 million euros and 7 million pounds sterling. The most significant media event of the year is the ARCO art fair (33), and a search of the newspaper archive demonstrates that coverage of this event is regular, with a focus on scandalous aspects.
- A relationship with politics (18) and power (25) is evident, and the use of the adjective “public” (32) is noteworthy. For instance, the word "freedom" is used only eight times, whereas the word "brand" is used 36 times. The use of the plural "women" (30) is to be preferred over the singular "woman" (13). In 2017, Barceló and Picasso were the most frequently cited artists, with a total of forty mentions (22 and 18, respectively).

It is logical to infer that the utilisation of automatic information analysis systems enables the management of significantly larger volumes of data. Employing the same programme structure in Python, a systematic check was conducted for the years 2016 and 2018 to ascertain the stability of the analysis across all its editions (morning, afternoon and evening). A total of approximately 100,000 posts on the home page were analysed, equating to 200,000 posts in total.

When the analysis was repeated for 2018, it was confirmed that the stability of the analysis was maintained, i.e. that the hypotheses were once again fulfilled and that this was a continuing trend. Concurrently, it was ascertained that the findings of the machine analysis are congruent with those previously articulated in the preceding section employing the manual Stempel method. This observation signifies that, henceforth, it would be considerably more rational to execute them automatically whenever feasible, in both instances. Subsequent to the implementation of the programme in 2018, and prior to the advent of the pandemic caused by the novel coronavirus known as SARS-CoV-2, the following results were obtained following the repetition of all analyses: A total of 129,534 posts were analysed, 7,688 of which were related to sports (5.9%), 10,330 to culture (8%), 2,957 to cinema (2.3%), 2,930 to art (2.3%) and 135 (less than 1%) to contemporary art. In summary, the findings of this study demonstrate that the medium's behaviour is consistent and reliable, thereby validating the hypotheses proposed.

4. Conclusions: A Comparison of Results Between the Stempel and Web Scraping Methods

This study corroborates the reliability of the findings derived from two quantitative analysis methodologies when employed on a specific case study, namely contemporary art news on ElPais.com over the course of a year. The year 2017 was considered pivotal as it marked the

centenary of the birth of the ready-made and was statistically prototypical as it preceded the emergence of the pandemic caused by the COVID-19 pandemic. Firstly, classic or manual sampling (Stempel) is carried out, with confirmation of data stability for the previous and subsequent years. Secondly, the analysis is refined with systematic sampling (2017, also comparing with the previous and subsequent years) using web-scraping techniques programmed in the Python programming language for the digital edition of the newspaper.

The findings of both methods were found to be consistent, although two drawbacks were identified from the outset: difficult access to newspaper archives and inconsistent use of tags in the categorisation of posts. Despite this, it has been widely confirmed that contemporary art is covered much less frequently than sport in the newspaper El País, and that its representation is often biased. The initial technique examined by El País is painting, and the notion of exhibition or show constitutes the majority of the content. That is to say, its approach to contemporary art can be characterised as conservative, insofar as it does not prioritise the creative process, the oeuvre of the artists, or the social significance of the art. El País's primary approach to contemporary art is as an "agenda", i.e. as a component of leisure or a strategy for allocating free time, in addition to the "lifestyle" focus.

The utilisation of the concept of "event" or focus on contemporary creations deemed to be of significant newsworthy interest is evident, with two annual moments of particular media significance. Firstly, the major annual event, which is covered by the ARCO fair, usually held in February, and for which the dominant trend in media coverage is that of scandal or absurdity, creating a framing that is repeated year after year at the commencement of the event with the same coverage. Secondly, the beginning of the season in September, which also attracts considerable media interest. In this analysis, other terms that recur in relation to the coverage of contemporary art include "politics", "power" and "public", as well as "brand".

In the context of digitalisation and automated systems that generate their own results, quantitative analysis in the field of art and social sciences is a powerful tool that can be used in two main ways. Firstly, it can be used as a method of content control, ensuring that important content is not overlooked within the medium itself. Secondly, it can be used as a tool for the external analysis of the media and its social impact, through independent observatories of information quality, and even as a preliminary step to trainable models.

Automation has been demonstrated to offer a more efficacious means of hypothesis formulation than the Stempel method, obviating the necessity for statistical analysis and the central limit theorem. However, it may be of interest to employ solely data from constructed weeks when optimised performance checks are required. The combination of new computational techniques with web scraping is pivotal in identifying trends and biases. However, it is essential that the media facilitate access to their content for research purposes, both quickly and automatically, in order to progress towards new Artificial Intelligence tools that perform these calculations and progressively refine their own models.

References

Bardin, L. (1986). *El análisis de contenido*. Akal.

Bauman, Z. (2007). *Vida de consumo*. Fondo de Cultura Económica de España.

Bauman, Z. (2018). *Modernidad líquida*. Fondo de Cultura Económica de España.

Berelson, B. (1952). *Content Analysis in Comunication Research*. Free Press.

Dans, E. (2012). Resistencia al cambio. In J.L. Orihuela (Ed.), *80 claves sobre el futuro del periodismo* (pp. 56-57). Anaya Multimedia.

Davis, F. J. & Turner, L. W. (1951). Sample Efficiency in Quantitative Newspaper Content Analysis. *Public Opinion Quartely*, *15*(4), 762-763. https://doi.org/10.1086/266358

Dimaggio, P., Nag, M. & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570-606. https://doi.org/10.1016/j.poetic.2013.08.004

Fernández-Costa, R. (2021). *Del escándalo al duelo: análisis de las tendencias informativas sobre el arte contemporáneo en la prensa online* [unpublished Doctoral Thesis]. Facultad de Ciencias de la Información de la Universidad Complutense de Madrid.

Fernández del Moral, J. (ed.) (2007). *El análisis de la información televisiva: hacia una medida de la calidad periodística*. Dosat.

Galtung, J. & Ruge, M. H. (1965). The Structure of Foreign News. *Journal of Peace Research*, *2*(1), 64-91. https://www.jstor.org/stable/4230111

Groys, B. (2014). *Volverse público. Las transformaciones del arte en el ágora contemporánea*. Caja Negra.

Guallar, J. (2012). *Las hemerotecas de la prensa digital. Análisis de diarios españoles* [Doctoral Thesis]. Universitat de Barcelona.

Guasch, A. M. (2007). *El arte último del siglo XX. Del posminimalismo a lo multicultural*. Alianza Editorial.

Hester J.B. & Dougall E.K. (2007). The Efficiency of Constructed Week Sampling for Content Analysis of Online News. *J&MC Quarterly*, *84*(4), 811- 824. https://doi.org/10.1177/107769900708400410

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison Wesley.

IAB (2017). *Estudio Anual de Medios de Comunicación Online.* https://iabspain.es/estudio/estudio-anual-de-redes-sociales-2017/

Jensen, K. B. (2014). *La comunicación y los medios. Metodología de investigación cualitativa y cuantitativa.* Fondo de Cultura Económica.

Jones, R. L., & Carter, R. E., Jr. (1959). Some procedures for estimating "news hole" in content analysis. *The Public Opinion Quarterly*, *23*(3), 399–403. https://www.jstor.org/stable/2746391

Krippendorff, K. (1990). *Metodología de análisis de contenido. Teoría y práctica*. Paidós.

Kristensen, N. N., & From, U. (2015). Cultural Journalism and Cultural Critique in a Changing Media Landscape. *Journalism Practice*, *9*(6), 760-772. https://doi.org/10.1080/17512786.2015.1051357

Lozano, J. & Martín, M. (Coords.) (2018). *Documentos del presente: una mirada semiótica*. Lengua de Trapo.

Luhmann, N. (2005). *El arte de la sociedad.* Editorial Herder y UIA.

Luhmann, N. (2000). *La realidad de los medios de masas.* Anthropos.

Python Software Foundation. *Python Language Reference, versión 3.7.4.* http://www.python.org

Purhonen, S., Heikkilä, R., Hazir, I. K., Lauronen, T., Fernández, C., Gronow, J. (2019). *Enter Culture, Exit Arts? The Transformation of Cultural Hierarchies in European Newspaper Culture Sections, 1960–2010.* Routledge.

Riffe, D., Aust, C. & Lacy, S. (1993). The Effectiveness of Random, Consecutive Day and Constructed Week Sampling in Newspaper Content Analysis. *Journalism & Mass Communication Quartlerly*, *70*(1), 133-139. https://doi.org/10.1177/107769909307000115

Rodríguez Pastoriza, F. (2006). *Periodismo cultural*. Síntesis.

RStudio Team (2020). *RStudio*: *Integrated Development for R. RStudio. PBC.* http://www.rstudio.com/

Stempel, G. H. (1952). Sample Size for Classifying Subject Matter in Dailies. *Journalism Quarterly*, *29*(3), 333-334.

Stempel, G. H. & Westley, B. H. (1989). *Research methods in mass communication*. Prentice Hall.

Tilles, D. (2016). The Use of Quantitative Analysis of Digitised Newspapers to Challenge Established Historical Narratives. *Roczniki Kulturoznawcze Journal*, *7*(1), 83-97. https://doi.org/10.18290/rkult.2016.7.1-4

UNESCO (2009). *Marco de estadísticas culturales (MEC) de la UNESCO 2009*. https://unesdoc.unesco.org/ark:/48223/pf0000191063