



# ARTE CONTEMPORÁNEO EN EL PERIÓDICO *EL PAÍS*: DOBLE ANÁLISIS DE LOS CONTENIDOS

## Comparación de resultados usando el método *Stempel* y el análisis computacional con programación en *Python*

RUBÉN FERNÁNDEZ-COSTA O'DOHERTY<sup>1</sup>, ELVIRA CALVO GUTIÉRREZ<sup>1</sup>, JOAQUÍN SOTELO GONZÁLEZ<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid, España

PALABRAS CLAVE	RESUMEN
Arte contemporáneo Periódico El País Stempel Cuantitativo Estudios automatizados Phyton Webscrapping Ciencia de datos	<i>La presente investigación confirma la coincidencia de un análisis doble repetido de los contenidos de arte contemporáneo en el periódico El País, usando en primer lugar la metodología Stempel y, en segundo, su automatización con técnicas de web scraping mediante programación en lenguaje Phyton. A la espera de sistemas completamente autónomos de IA que hagan posible un seguimiento instantáneo de los contenidos de un sistema informativo, ambos enfoques confirman que –cuando se cumple un siglo del tradicionalmente aceptado como nacimiento del arte contemporáneo (Duchamp, 2017)– existe una escasa representación del arte actual en un medio de comunicación icónico y referente en España, como es el analizado, para un año estadísticamente estándar, frente, por ejemplo, a las informaciones de la sección de Deportes y otras categorías culturales, como Cine y Series, de ese mismo medio. Además, se detecta un uso de vocabulario hiperbólico y materialista en la representación periodística del arte actual.</i>

RECIBIDO: 27 / 02 / 2025

ACEPTADO: 10 / 06 / 2025

## 1. Introducción

La aplicación de técnicas cuantitativas a los ámbitos artísticos y comunicativos ha ido incorporándose gradualmente a lo largo de la segunda mitad del siglo XX hasta llegar a consolidarse su uso habitual en las Ciencias Sociales y las Humanidades (Bardin, 1986; Berelson, 1952; Holsti, 1969; Krippendorff, 1990). En memoria del profesor fallecido Dr. Jorge Lozano —catedrático de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid— podemos afirmar incluso que vivimos una «nueva Tercera etapa», en la que, gracias a la analítica cultural a gran escala (Lozano y Martín, 2018), por primera vez es posible realizar análisis cuantitativos de grandes volúmenes de datos que nos permiten obtener resultados sobre los que reflexionar (DiMaggio et al., 2013; Kristensen y From, 2015; Tilles, 2016). Para esta «Tercera etapa», el presente artículo pone el foco en dos consideraciones previas a los análisis cuantitativos de sistemas informativos. En primer lugar, la necesidad de contar con un conocimiento profundo y experto en la materia —en este caso, sobre el complejo ámbito del arte contemporáneo y sobre su representación en un escenario de comunicación cambiante— y un compromiso ético claro: que las jerarquías culturales y narrativas históricas son «multicasuales» y que en la idiosincrasia propia de un sector existen múltiples implicaciones del lenguaje (Bauman, 2007; Bauman, 2018; Purhonen et al., 2019). El presente estudio no obvia este hecho, sino que lo subraya. En segundo lugar, es imprescindible confirmar no sólo la viabilidad de los *Software Studies* en cuanto a la accesibilidad de los datos, sino respecto a la calidad y robustez de las métricas, haciendo uso de la estadística para refrendar las hipótesis con una posterior revisión cualitativa.

Es ampliamente aceptado que los medios de comunicación, y específicamente los periódicos, contribuyen a modelar la opinión pública por medio de la elección de temas —*agenda setting*—, ángulos de noticiabilidad y encuadres informativos —*news values / framing*—, siendo denominados en sí mismos como «instituciones con las que se piensa» (Jensen, 2014), puesto que intervienen en los intercambios entre diferentes sistemas sociales (Luhmann, 2005; Luhmann, 2000). En España, su influencia es clave: según el Estudio Anual de Medios de Comunicación Online (IAB, 2017), dentro de los medios de comunicación, los que tienen una mayor penetración son los periódicos online, seguidos de los portales temáticos. Además, las webs de información cuentan con un mayor nivel de confianza y credibilidad respecto a la radio, a la televisión y a los portales temáticos.

Hoy en día, los «diarios» han dejado de ser «periódicos» y han pasado a ser «continuos», permanentemente accesibles, consumidos en dispositivos móviles y replicados en redes sociales (Dans, 2012, p. 57), por lo que «el periódico online se alza como el medio digital en el que los españoles, especialmente [...] individuos a partir de 30 años, consultan más veces a la semana» (IAB, 2017, p. 14). En España, *elpais.com* era, según la medición de *Comscore* para 2017, el medio online más leído (2.834.000 visitantes únicos), además de un medio de gran influencia y tradición.

En lo que respecta al arte contemporáneo, en el año 2017 se cumplió un siglo desde la que se considera según el relato occidental como la primera obra contemporánea (el *ready-made* «El Urinario», de Marcel Duchamp, realizado en 1917), obra que simbólicamente inauguró una nueva época caracterizada por una gran libertad comunicativo-artística, colocando al artista en el centro de la decisión de qué es y qué no es arte, poniendo el foco en el emisor y no en la dificultad de realización o la temática de la obra. Este giro copernicano de postura provocó, según Anna María Guasch (2017), el inicio de un completo desajuste entre los lectores de los medios de masas y la creación más radicalmente contemporánea, puesto que «el público no reconoce como suyo el arte de su época, actitud de la que deriva el rechazo o, cuando no, una cierta acusación de ininteligibilidad, opacidad discursiva y dispersión» (Guasch, 2007, p. 17). En resumen, el año 2017 es clave en este estudio porque a) corresponde al momento exacto en que se cumplen 100 años del nacimiento del arte contemporáneo (1917); b) es un año estadísticamente típico por ser previo a hiperacontecimientos o *booms* informativos que suelen provocar fuerte recesión expositiva de determinados temas, como sucedió durante la pandemia por COVID-19, acontecimiento que eclipsó informativamente muchos otros temas (los años 2019-2022 son considerados por los expertos como años de «recuperación» de visitantes a las actividades artísticas y culturales), c) porque no corresponde a una gran efeméride o celebración nacional del arte contemporáneo (no obstante, se detectarán los 'ochenta años' del nacimiento de Picasso); d) en el presente estudio se confirman los resultados con el año previo y posterior, también considerados como 'típicos'.

El presente trabajo está en línea con la investigación desarrollada en la tesis doctoral *Del escándalo al duelo: un análisis de las tendencias informativas sobre el arte contemporáneo en la prensa online*

(Fernández-Costa, 2020), y propone un ejercicio doble: comprobar si la propuesta comunicativa de contemporaneidad sigue vigente en un medio líder por medio de dos análisis, un análisis manual y uno automático consecutivos, partiendo de las siguientes hipótesis: H1) bajo porcentaje de contenidos de Arte contemporáneo respecto a otros, como de Deporte; H2) máxima importancia al cine y las series en la sección de Cultura y H3) uso y abuso de ciertas palabras en el modo de cubrir periodísticamente la creación artística actual. Se pretende localizar también en estos contenidos sesgos de duelo — desafección con la creación— y fórmulas comunicativas usadas por los medios de comunicación respecto a esta situación.

El periódico *El País* (Grupo PRISA) se publicó por primera vez en el año 1976 y constituye no sólo el diario generalista de pago con mayor difusión en España en el momento de análisis (categoría que varía según el mes), sino que, además, se trata de un medio con una larga tradición cultural. En un primer paso, se realiza para este periódico el análisis de una gran cantidad de datos recopilados manualmente desde la hemeroteca de *El País* a partir del método Stempel y los conceptos de «semanas construidas» y «carácter cíclico anual». Como afirma Javier Guallar (2012) en su tesis doctoral *Las hemerotecas de la prensa digital. Análisis de diarios españoles*, es importante para los investigadores en Ciencias Sociales que los medios de comunicación no dificulten las consultas de sus hemerotecas.

En un segundo apartado posterior, se realiza un análisis automático de las mismas cuestiones, usando técnicas de *web scraping* basadas en el *taggeado* de artículos por parte del medio. Se trata de dos pasos previos de suma importancia para comprender adecuadamente el procesado y monitorización de los medios que puede ser de ayuda en un estadio previo a la llegada de programas completamente autónomos —e incluso de caja negra— de Inteligencia Artificial (IA, ver *Harvard Consilience*), pudiendo ser de utilidad para identificar sesgos y alucinaciones en los análisis y que arrojen luz sobre los resultados obtenidos.

## 2. Método manual [Stempel] de análisis cuantitativo: monitorización de los contenidos de «arte contemporáneo» en la edición online del periódico *El País*

El análisis cuantitativo de un medio de comunicación periódico se realiza habitualmente en el ámbito científico usando el método de la «semana construida» (Krippendorff, 1990; Stempel y Westley, 1989). Una semana construida consiste en un lunes aleatorio, seguido de un martes aleatorio y así repetidamente hasta llegar a un domingo aleatorio y completar los siete días elegidos de todas las fechas posibles. Fue Stempel (1952) el primero en determinar que dos semanas «construidas» eran suficientes para representar el comportamiento de un medio de comunicación durante un año completo. Posteriormente, Davis y Turner (1951), Jones & Carter (1959) y Riffe et al. (1993) demostraron que dos semanas «construidas» son, de hecho, suficientes, y Hester y Dougall (2007) añadieron que se debían duplicar al menos los datos para caracterizar un medio online.

El presente trabajo estudia, como unidades de análisis, los textos informativos que aparecen publicados en la web del medio objeto de investigación, en adelante denominados *posts* o noticias. En la hemeroteca de *El País*, como confirma el propio medio, desde el día 7 de febrero del año 2012, los contenidos digitales de esta cabecera coinciden con los del papel. La ventana del presente análisis corresponde inicialmente a un año, un ciclo completo de los eventos culturales y artísticos habituales, ayudando así a detectar malas prácticas y buenas tendencias informativas.

Se han analizado manualmente un total 1.271 *posts*, considerándose las tres versiones diarias de la *home* o portada del diario (mañana, tarde y noche): sin el método Stempel, el volumen de datos no sería asumible, puesto que 100 *posts* nuevos diarios corresponderían a 36.500 *posts* nuevos anuales, número que se triplicaría por la existencia de tres ediciones diarias. En este caso, usando un muestreo estratificado y las funciones de *Excel Randbetween* y *Fecha* se han obtenido los listados de fechas de este muestreo manual, dos para cada día de la semana y por cada semestre (dos fechas de lunes, dos fechas de martes, dos fechas de miércoles...) del 2017, de enero a diciembre. De este modo, se llega a 1.271 *posts* (sin incluir los *posts* no artísticos de las ediciones de la tarde y noche), de los cuales se localizan 402 relativos a cultura y 43 a arte contemporáneo.

Para clasificar las temáticas de los *posts* de *El País*, se utilizan como base las categorías de *Dominios culturales* (UNESCO, 2009) y, cuando surgen dudas en la clasificación, se toma siempre la decisión contra la hipótesis para asegurar el comportamiento estadístico. Así, se definen las categorías culturales de «arte contemporáneo y actual», «arte clásico, patrimonial o funcional (arquitectura y diseño)», «artes escénicas», «cine y series», «literatura actual», «literatura clásica», «música actual» y «música clásica».

Metodológicamente, el investigador contabiliza los *posts* hasta tres veces en cada medida, obviando las publicidades y *banners*, y los apartados de «más leídos».

La ficha de análisis y codificación, de elaboración propia, se ha diseñado basándonos en los trabajos de Galtung y Ruge (1965) y Fernández del Moral (2007), con un total de 20 campos (algunos dependientes de otros), con variables y valores de distinta naturaleza, incluyendo: «Día de la semana (L-D)», «Semestre (1 ó 2)», «Semana construida (1-5)», «Edición (mañana, tarde o noche)», «Número total de noticias en *Home*», «Total de noticias sobre Arte actual», «Noticia», «Titular», «Enfoque», «Texto completo», «Número de noticias Cine y Series, con sus titulares», «Número de noticias Teatro y Artes escénicas (inc. ópera), con sus titulares», «Número de noticias de Música Clásica, con sus titulares», «Número de noticias Música no Clásica, con sus titulares», «Número de noticias Literatura Clásica, con sus titulares», «Número de noticias Literatura Actual, con sus titulares», «Número de noticias sección de Cultura (automática, atendiendo a su clasificación web)», «Número de otras noticias relacionadas con el arte no contemporáneo (arte clásico, patrimonio, arquitectura,...), con sus titulares», «Número Total noticias culturales a criterio observador del investigador, analizadas cada día (suma de todas las categorías)» y «Número de noticias sección Deporte».

## 2.1. Resultados del análisis manual [Stempel]

Los resultados que se obtienen en este apartado corresponden a una ventana de observación de un año completo, período natural de estudio, porque la mayoría de las noticias y eventos son cíclicos con esa periodicidad. En media, el número total de noticias nuevas que publican online cada día en el periódico *El País* es de 90,79 piezas (se comprueba que otros medios online también realizan una media cercana a los cien *posts*). La mediana y dispersión típica confirman que los resultados se distribuyen de manera estadísticamente habitual, en forma de campana gaussiana y simétrica, con una mediana de 88 y una desviación típica de 8,37 respecto a la medición de nuevos *posts* por día. El proceso de conteo se realiza de forma manual y se considera que el error humano introducido en la desviación estándar del proceso de conteo es despreciable estadísticamente por tratarse, las mediciones sucesivas, de un error aleatorio que se considera que se ha de repartir por exceso y/o por defecto.

En relación a los resultados de la categoría definida en la ficha de elaboración propia como 'Cultura', cada día se publican de media 28,71 *posts*, con una mediana de 29 y una desviación típica de 6,18. En relación a las noticias de Arte Contemporáneo en *El País*, la media se sitúa en 3,07 noticias, con una mediana de 3 y una desviación típica de 1,54. Para el resto de categorías culturales, se obtienen los siguientes resultados: 10,29 piezas en «Cine y Series» (media 8,5; mediana 5,85) y 6,07 en 'Literatura' (media 6,5; mediana 5,85). Para la categoría diferencial respecto de la hipótesis sobre el bajo porcentaje de contenidos de Arte contemporáneo respecto a otros (H1), se mide también para esta investigación la categoría de «Deporte» y se obtienen para ella 5,43 piezas diarias de contenido (media 5; mediana 0,65).

Al publicarse diariamente siempre alrededor de 100 contenidos nuevos por día, los porcentajes obtenidos son similares a los totales, como se podrá observar a continuación. En proporciones, se obtienen unos porcentajes de «Cultura» de 31,70% (con mediana de 32,48% y dispersión típica de 6,94%); «Arte contemporáneo» de 3,40% (mediana 3,22% y dispersión de 1,76%); «Cine y series» de 11,35% (9,09% y 6,69%) y «Literatura» de 6,72% (7,02% y 2,56%). Respecto a la categoría diferencial de análisis de «Deporte», se obtienen los resultados de 16,40%, 15,80% y 0,94%.

## 2.2. Conclusiones de los resultados del análisis manual [Stempel]

Como se ha comentado, apoyándonos en el teorema central del límite, y dado que tenemos un tamaño de muestras suficientemente grande, podemos asumir que las proporciones o medias se distribuyen de forma normal o gaussiana, y calcular intervalos de confianza para contrastar las hipótesis.

En primer lugar, respecto a la primera de nuestras hipótesis planteadas párrafos atrás (H1) sobre el bajo porcentaje de contenidos referidos a arte contemporáneo y creación artística actual en la *home* de *El País*, especialmente, respecto a otras secciones, como Deporte, se obtiene que la media de noticias de arte contemporáneo se sitúa en 3,07 noticias, con una mediana de 3 y una desviación típica de 1,54; es decir, en un curva gaussiana y aplicando el Teorema Central del Límite, con un margen del 90% de confianza [ $P \pm 1.65 \cdot \sqrt{((p(1-p))/98)}$ ], se obtienen los límites inferior de -4,59% —que se puede colocar a cero— y superior de 11,40%. Por tanto, la media («p») calculada oscila al 90% de confianza entre 0 y 0,114, es decir, la proporción media de noticias relacionadas con la creación artística actual es menor

que el 11%, siempre inferior a las medias porcentuales del resto de las temáticas, incluyendo a la de «Deporte», que oscila entre 15,46% y 17,34%, con lo cual la primera hipótesis (H1) se cumple.

En segundo lugar, respecto a nuestra segunda hipótesis (H2) sobre que para la sección de cultura se da máxima importancia al cine y las series, según los datos, se comprueba que, en términos absolutos, el contenido con mayor número de entradas –en media– en *El País* es el de «Cine/Series» (seguido de «Literatura moderna»). No obstante, la distribución de proporciones de noticias de cine aparece situada entre las categorías de arte y deporte, lo cual no permitiría concluir que sea así siempre y para cualquier día, sin embargo, aparece como altamente habitual.

En tercer y último lugar, respecto a la hipótesis (H3) de uso abusivo de ciertas palabras en el modo de cubrir informativamente la creación artística actual, se realizan dos pasos previos: para todos los textos recopilados se eliminan las *stop words* y se realiza una comparación con los términos de frecuencias típicas de aparición de palabras del idioma castellano según el *Corpus de Referencia del Español Actual* (CREA, accesible online a través de la Real Academia Española, RAE). Para el caso de *El País*, se trata de 33.905 palabras, con cuyo análisis relacional se determinan las siguientes características:

- Enfoque hiperbólico: uso repetido de la palabra «más» y de la función hiperbólica. Si en el corpus CREA de la RAE la palabra «más» estaría en el puesto 23, aquí encabeza el ranking. Se busca captar la atención con estructuras del lenguaje basadas en el adverbio, como «la más», «el más» o «lo más». Las palabras «vida» y «mundo» son los dos primeros términos abstractos de este análisis manual y la palabra «vez» también determina un planteamiento iterativo.
- Narrativa institucional: utilización de las palabras «exposición», «museos», «sala» y «obra». En este caso, el medio analizado decide incluir un contenido porque encaja dentro de una «percha», es decir, con la visión de «agenda» que permitirá a quien visite un centro de arte realizar una actividad de ocio que, a menudo, se consume como un producto más de la industria cultural. Como señala Groys, «los museos se han vuelto los sitios de las exhibiciones temporales más que los espacios para las colecciones permanentes» (Groys, 2014, p. 88).
- Planteamiento centralista: en su conjunto, la muestra da evidencias de centralista (Madrid-Barcelona) y destaca eminentemente a artistas masculinos. «Picasso» aparece como el primer nombre propio de la lista, en lo que seguro influye el 80 aniversario del Guernica.

### 3. Automatización del análisis cuantitativo usando *web scraping*: contenidos de «arte contemporáneo» en la edición online del periódico *El País*

Si bien un estudio de carácter manual o artesanal cuenta con la observación directa del investigador, pueden obtenerse resultados de mucho mayor volumen cuantitativo usando técnicas de *web scraping* y programación en *Python*, en la categoría de lo que actualmente se denomina *data science* o *software studies*.

El *software* utilizado en esta investigación ha sido Python 3.7.4, un lenguaje de programación interpretado, multiparadigma y orientado a objetos, de código abierto y gratuito, con librerías, fácil de depurar y flexible (crea un entorno fácilmente *debuggeable*). En este caso, se han utilizado las librerías *Pandas* y *BeautifulSoup*, las funciones generalistas *urllib* y *request*, cuyo uso conjunto permite realizar *web scraping* y *crawling*, así como el IDE —entorno de desarrollo integrado— *Pycharm* para desarrollar y probar el código previamente a su ejecución, *Atom* y *Sublime*, entre otros soportes. Se omiten las informaciones relativas al procedimiento técnico y el código generado de análisis, que siendo reutilizable para estas fechas, se pone bajo demanda por petición directa a disposición de la comunidad académica.

Es clave la elección de los *tags* que el programa diseñado debe rastrear. Para este análisis, se confirma que el idóneo es «Arte contemporáneo» del siguiente listado obtenido: ART, Artà, Artana, Art Barcelona, Art Basel, Art Blakey, Arte, Arte Abstracto, Arteaga, Arte antiguo, arteBa, Arte barroco, Arte bizantino, Arte conceptual, Arte contemporáneo, Arte degenerado, Arte digital, Arte egipcio, Artefactos caseros, Arte flamenco, Arte gótico, Arte griego, Arte ibérico, Arte islámico, Arteixo, Arteixo Telecom, Arte medieval, Arte mesopotámico, Artemio, Artemio Baigorri, Artemio Cuenca, Artemio Precioso, Artemi Rallo, Artemi Rallo Lombarte, Artemisia Gentileschi, Arte mudéjar, Arte Multimedia, Artenara, Artenbrut, Arte neoclásico, Arte pop, Arte povera, Arte prehistórico, Arte prerrománico, Arte religioso, Arte renacentista, Arte románico, Arte romano, Artés, Artesanía, Artesanos Camiseros, Artes

decorativas, Artes escénicas, Artes gráficas, Arte siglo XX, Artes ilegales, Artes marciales, Artespaña, Artes plásticas, Arte TV, Arte urbano, Artez, Art Futura.

Como sucede habitualmente, el medio analizado tiene una serie de características que determinan cómo se lleva a cabo su análisis y se depende de la consistencia de la construcción del conjunto. Los datos obtenidos en formato *csv* son transformados en tablas de excel y se realizan resúmenes y gráficos con el programa *R* y el entorno de desarrollo de *RStudio*. Optamos, como se explica más adelante, por el análisis de todas las portadas digitales del periódico *El País* a lo largo de un año completo. A continuación, se confirma la estabilidad de los resultados para los años anterior y posterior.

### 3.1. Resultados método web scraping

Usando *web scraping* obtenemos automáticamente una tabla que representa todo el año. Dado que el periódico *El País* tiene tres versiones diarias para mañana, tarde y noche (cuyas direcciones web concluyen en /m, /t y /n), la lista que genera el programa en primer lugar contiene 365x3 líneas, es decir, 1.095 «fechas» para un año.

El programa, al ser ejecutado, genera 7.793 *links* fallidos para un total de 108.547 (pierde sólo el 7,7% de los *links* en *home*), en su mayoría, se trata de *links* ya no existentes, sin texto o con contenido de vídeo, o *links* perdidos con «Error 403». En relación a los 108.547 *posts* que sí localiza, el programa de *web scraping* desarrollado encuentra 9.238 *posts* taggeados con la palabra «cultura» (8,5%) y 6.690 con la palabra «deportes» (6,2%).

Se confirma que los *posts* que se encuentran cada día en cada una de las versiones sucesivas de la *home* de *El País* (3 cada día) son, por configuración de la rejilla, destacados a diario, por lo cual la medida de *home* caracteriza el comportamiento del medio. Asimismo, confirma que el *web scraping* vuelca todos los datos deseados: 108.547 *posts* totales analizados, que corresponden a 6.690 de «Deportes», 9.238 de «Cultura», 2.649 de «Cine», 9.238 de «Arte» y 143 de «Arte contemporáneo». En porcentajes, las categorías se distribuyen como 6,2% de «Deportes», 8,5% de «Cultura», 2,4% de «Cine y series», 1,9% de «Arte» y menos del 1% de «Arte contemporáneo».

### 3.2. Conclusiones del análisis automático: web scraping

Respecto a las hipótesis de la investigación, el programa localiza 2.064 *posts* que contienen el tag «arte», es decir, cerca del 2%; 143 con el tag «arte contemporáneo», lo cual nos indica un porcentaje del 0,13%. Incluso considerando (caso peor) incluir en el conteo los *posts* otros tags como «arte del siglo XX» (110) y «arte urbano» (62), los dos únicos tags significativos del listado del diccionario de tags de *El País*, el porcentaje seguiría siendo inferior al 1% —0,29%— de los destacados en *home*, por lo que las hipótesis se cumplen.

En una última comprobación, incluso tomando la decisión de incluir el restante de *posts* taggeados en la categoría (90+172), obtenemos igualmente un porcentaje inferior al 1%, es decir, se podría señalar que la cobertura del «arte contemporáneo» no es un objetivo prioritario del medio y es sustancialmente inferior al resto de las categorías definidas (7,02% de «deporte» frente a 0,24% de «arte contemporáneo»).

En este porcentaje están incluidos, como decisión en contra de la hipótesis, los enlaces que siguen siendo válidos en el momento de la muestra a *El País Semanal*, *The Huffington Post* y resto de publicaciones embebidas en esta *cultura mosaico* de la que en su día habló el filósofo francés Abraham Moles. En este caso, *boxplots* e histogramas confirman todas las medidas de los momentos del día, es decir, 3 (mañana, tarde y noche) en una sucesión temporal de 1.095 muestras temporales anuales, que pueden contener *posts* repetidos (de nuevo, una decisión en contra de la búsqueda), pero las hipótesis se cumplen en todos los casos.

A continuación, por medio del uso del programa *RStudio*, se puede realizar un análisis completo de las tendencias encontradas en el volumen total de datos recogido para el periódico *El País* durante todas las ediciones del año 2017. No se incluyen por simplificación, pero están disponibles para la comunidad investigadora los gráficos de análisis de este apartado. Se comprueban las siguientes conclusiones:

- El número de *posts* diarios en *home* es de alrededor de 100, que corresponde aproximadamente al nuevo contenido generado diariamente.
- Los *posts* taggeados como «arte contemporáneo» son destacados en *home* (portada digital) en escasas ocasiones por *El País*. En todo caso, para *El País*, el número de *posts* con el tag «cultura», en sentido amplio, que se destacan es de un orden de magnitud comparable al de «deporte».

En relación a las dos primeras hipótesis (H1 y H2), los porcentajes de *posts* de «cultura» presentan un comportamiento variable en función de la fecha, según lo esperado. Por ejemplo, el máximo absoluto de contenidos taggeados como «cultura» correspondiente al 26 de noviembre de 2017 lleva en portada una cobertura especial de la Feria del libro de Guadalajara, evento en el que *El País* fue *partner*. En los meses de noviembre y diciembre del periodo analizado, se observa asimismo que se han colocado más contenidos culturales generales en portada. Se observa un comportamiento estacional (coincidiendo con la feria ARCO) para la categoría «Arte contemporáneo» (los 6 primeros máximos de la tabla son en febrero, apareciendo también marcado el inicio de temporada en septiembre). Sin embargo, en la comparativa Cultura/Arte contemporáneo se confirma un porcentaje muy menor de la creación artística contemporánea respecto de otras propuestas culturales.

Para resolver la tercera hipótesis (H3), analizamos un listado limpio consecutivo de los titulares de *posts* en relación al arte contemporáneo durante el año 2017 para una descripción de la «construcción del discurso» con «sesgos» o «perchas». El sesgo principal encontrado es el de «reducción al absurdo», muy similar a la idea de «escándalo», cercano a la mitad de los titulares.

**Tabla 1.** Titulares localizados por el tag «arte contemporáneo» en *El País* 2017

Entre Lenin sin cuerpo y Franco sin cabeza
La misteriosa escena de decapitación de la que habla Bruselas
Manual de acupuntura
«Resetear» el expresionismo abstracto
Ese camino que conduce a una estrella
Argentina se sacude la nostalgia en Arco 2017
Inquietante extrañeza de David Lynch
Elogio del corta y pega
La vertiginosa crítica institucional
En el mismo barco
El muro que se mueve
Cuando Warhol visitó España por última vez
Arco hincha el optimismo
Los Reyes y el presidente argentino inauguran Arco 2017
¿Hasta un niño podría crear arte contemporáneo? Los pequeños de Arco 2017 responden
Isa Calderón y Amarna Miller hacen un examen de porno en ARCO
Julian Rosefeldt: «Vivimos en una época de masturbación cultural»
Sevilla estrena CaixaForum
Eso era todo
Cuando el rumor se materializa y aparece Obama en un museo
La «banlieue» apuesta por el arte para contrarrestar la marginación
Poetas en el museo
Bajo los adoquines, el museo
Martí Fluxá sustituye a Guillermo de la Dehesa en el patronato del Reina Sofía
Versalles en el «down-town»
Las cenizas de Luis Barragán llegan a México convertidas en diamante
Miquel Barceló instala en Salamanca su Arca de Noé
Otra forma de ser políticos
Bienal de Venecia: ¡que empiecen los juegos del arte!
La Bienal En Marche (LBM)
Basquiat bate con 99 millones de euros el récord en subasta para un artista americano
Arte en papel
Como en 'Sherlock Holmes', todos mentían
¿Una hija del gran masturbador?
Un tesoro artístico se dirige a España
Arroyo asciende al Olimpo del arte de la Costa Azul
Sonia y el futuro absoluto
El gran centro de arte que emerge de la decadencia industrial de Valencia
Bleda y Rosa: «Miramos hacia geografías menores, frágiles»
¿Cuánto cuestan «Las meninas»?

Prefirió no hacerlo
Richard Serra: «El mejor arte es intrínsecamente inútil»
Pantallas en fuga
Pintura con pies de hormigón
Cuando la musa es un emoji
La verdad como ensayo
San Sebastián: arte, comida, playa y controversia
Instagrameables o el asalto de las imágenes
Doris Salcedo: «Lo difícil es lograr una imagen invisible, una iconografía sutil»
Cuando el sexo era demasiado explícito
Y Masotta cometió un «happening»
Kusama, la reina nipona de los lunares
«Pintar en árabe no es hostil, es poesía, y eso llama al diálogo»
Marruecos se impregna de arte español
José Luis Alexanco: «El arte se ha convertido en un espectáculo»
Cuerpos desnudos y feminismo, así es el universo de María Hesse
Liberado un artista belga tras 19 días encadenado a un bloque de mármol
Eduardo Arroyo, siglo XXI
Yoko Ono instala una escalera para subir al cielo de Córdoba

Fuente: Elaboración propia con *web scraping* en *Python* (2024)

Para el conjunto de los 76 titulares completos (57.061 palabras) sobre arte contemporáneo que han ocupado las portadas del periódico a lo largo del año completo, se realiza una operación de compactado (eliminando las *stop words*) y se calculan las frecuencias absolutas y relativas de aparición. Se obtienen las siguientes conclusiones:

- Planteamiento «conservador»: la primera técnica reseñada es la pintura con *fact* (frecuencia absoluta de aparición de una palabra en un texto o conjunto) de 39 (*pintor*, 18), seguida de la *escultura* (21), la *instalación* (19) y la *fotografía* (17). Se le da asimismo máxima importancia al concepto de *muestra* o *exposición* (53+78), muy por encima de *colección permanente* (39), en máxima correlación con las etiquetas de «agenda» y «tiempo libre». Se relata sobre todo lo que sucede en el *museo* (112) o *museos* (46) por encima de *espacio* (44) o *espacios* (15) y con una preferencia centralista, con el *Reina Sofía* (23) como lugar preferido y *Madrid* (51) muy por encima de *Barcelona* (16).
- Construcción lingüística que busca la «intensificación hiperbólica», con fraseología comparativa usando el adverbio *más* (292), pero también con *nueva* (71), *mucho* (49) y *hasta* (108). También *nunca* (28) y *siempre* (49). En relación al argumento escándalo, a pesar de la disparidad del léxico de este ámbito, resalta el uso de los términos *sexual* (11), *sexo* (8) y *sexualidad* (9).
  - En relación al argumento comercial, *mercado* tiene una *fact* de 23, *millones* tiene una de casi el doble (41), además de *euros* (32) y *dinero* (7). El gran evento mediático del año es la feria *ARCO* (33) y se puede comprobar en hemeroteca que la cobertura de este evento es habitual, buscando propuestas con el sesgo «escándalo».
- Existe una relación con *política* (18) y *poder* (25) y resalta el uso del adjetivo *público* (32). Como comparativa, se puede observar, por ejemplo, que la palabra *libertad* sólo se localiza 8 veces, mientras que la palabra *marca* se encuentra en 36 ocasiones. Se detecta el uso del plural *mujeres* (30), preferido al singular *mujer* (13). En 2017, Barceló y Picasso son los artistas más mencionados, sumando ambos cuarenta menciones (22 y 18, respectivamente).

Como es lógico, el uso de sistemas automáticos de análisis de la información permite gestionar volúmenes de datos mucho mayores: usando la misma estructura de programa en *Python*, realizamos a continuación una comprobación sistemática para los años 2016 y 2018 de cara a confirmar la estabilidad del análisis, en todas sus ediciones (mañana, tarde y noche), analizando un volumen de alrededor de 100.000 *posts* en *home*, es decir, 200.000 *posts* estudiados.

Al realizar de nuevo el análisis para el año 2018, se confirma que se mantiene la estabilidad del análisis, es decir, que las hipótesis se cumplen de nuevo y que se trata de una tendencia que se mantiene. En paralelo, se confirma que las conclusiones del análisis maquinal coinciden con las anunciadas por el apartado anterior que usa el método manual Stempel, por lo cual tendría en un futuro mucho más sentido realizarlas de forma automática siempre que esto fuera posible, en ambos casos. Repitiendo todos los análisis con el programa realizado para el año 2018, también previo a la pandemia por COVID-

19, se obtienen los siguientes resultados: 129.534 *posts* totales analizados, 7.688 de Deportes (5,9%), 10.330 de Cultura (8%), 2.957 de Cine (2,3%), 2.930 de Arte (2,3%) y 135 (inferior a 1%) de Arte Contemporáneo. Es decir, se confirma la estabilidad de comportamiento del medio y todas las hipótesis.

#### 4. Conclusiones: comparativa de resultados entre los métodos Stempel y *web scraping*

El presente trabajo confirma la coincidencia de los resultados de dos maneras de análisis cuantitativo para un caso de estudio concreto, las noticias de arte contemporáneo en *ElPais.com* durante un año completo, el año 2017, clave al cumplirse un siglo del nacimiento del *ready-made* y tratarse de un año estadísticamente prototípico por ser previo a la pandemia por COVID-19. En primer lugar, se realiza el muestreo clásico o manual (Stempel), con una confirmación de la estabilidad de los datos para los años previo (2016) y posterior (2018). En segundo lugar, se refina el análisis con un muestreo sistemático (2017, comparando asimismo con los años previo y posterior) usando técnicas de *web scraping* programadas en el lenguaje de programación *Python* para la edición digital de dicho periódico.

Se comprueba que los resultados de ambos métodos son consistentes, si bien se localizan dos inconvenientes desde un principio: el difícil acceso a las hemerotecas y el uso inconsistente de las etiquetas en la categorización de los *posts*. A pesar de ello, en ambos casos, se confirma ampliamente que el arte contemporáneo aparece reseñado en el periódico *El País* de manera mucho menos frecuente que el deporte y que su representación es a menudo sesgada. La primera técnica artística reseñada por *El País* es la pintura, y el concepto de exposición o muestra articula la mayoría de los contenidos, es decir, su aproximación al arte contemporáneo puede ser calificada de conservadora, puesto que no se centra en la creación en proceso, ni en el trabajo de los artistas, ni en la importancia social del mismo. El planteamiento principal de *El País* para el arte actual es como «agenda», es decir, como elemento de ocio o plan para el tiempo libre, además del enfoque de «estilo de vida».

Destaca el uso del concepto de «evento» o percha necesaria para la noticiabilidad de la creación actual, existiendo dos momentos anuales de mayor impacto mediático: 1) el gran evento anual de la feria *ARCO* —habitualmente durante el mes de febrero—, para el que la tendencia dominante de cobertura es la del *escándalo* o *reducción al absurdo*, creando un *framing* que se repite año tras año al inicio de ese evento con la misma cobertura y 2) el momento de mayor carácter mediático es el arranque de temporada en el mes de septiembre. Otros términos que se localizan recurrentes en relación con la cobertura del arte contemporáneo en este análisis son los de «política», «poder» y «público», así como el de «marca».

En relación a la utilidad de los análisis cuantitativos en el ámbito del arte y las Ciencias Sociales, en un momento de creciente digitalización y sistemas automáticos que presentan sus propios resultados, los métodos de análisis cuantitativo de los sistemas de información constituyen una herramienta de gran potencia que puede ser utilizada de dos modos principalmente: como método propio de control de contenidos, para no olvidar la importancia de algunos de ellos desde el propio medio; o como una herramienta útil para el análisis externo de los medios de comunicación y su impacto social, a través de observatorios independientes de la calidad de la información e incluso como un paso previo a los modelos entrenables de inteligencia artificial.

La automatización revela de forma más clara las hipótesis que el método Stempel, sin hacer uso de la estadística y el teorema central del límite, pero puede ser interesante usar sólo los datos de semanas construidas cuando se quieran realizar comprobaciones optimizadas de funcionamiento. Las nuevas técnicas computacionales unidas al *web scraping* son claves para localizar tendencias y sesgos, siempre que los medios permitan acceder a sus contenidos en términos investigadores, de una manera rápida y automática, en el camino hacia nuevas herramientas de inteligencia artificial que realicen estos cálculos y refinen progresivamente sus propios modelos.

## Referencias

- 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 Communication Research*. Free Press.
- Dans, E. (2012). Resistencia al cambio. En 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 Quarterly*, 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/423011>
- 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 Quarterly*, 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>