

**Cómo citar este artículo en bibliografías / Referencia**

R Sánchez Sabaté, C del Valle, M Mensa (2019): “Método para la construcción de grandes corpus temáticos de noticias de prensa digital. Hacia un corpus sobre el hecho alimentario”. *Revista Latina de Comunicación Social*, 74, pp. 594 a 617.

<http://www.revistalatinacs.org/074paper/1347/30es.html>

DOI: [10.4185/RLCS-2019-1347](https://doi.org/10.4185/RLCS-2019-1347)

# Método para la construcción de grandes corpus temáticos de noticias de prensa digital. Hacia un corpus sobre el hecho alimentario

Method for the construction of large thematic corpora of online news articles. Towards a corpus of food-related news

**Rubén Sánchez Sabaté** [CV] [ [ORCID](https://orcid.org/0000-0003-4717-2259)] <https://orcid.org/0000-0003-4717-2259> [ [GS](#)]

<https://scholar.google.com/citations?user=X7htsn4AAAAJ&hl=es&oi=ao>

Universidad de La Frontera, Núcleo Científico Tecnológico en Ciencias Sociales y Humanidades, Centro de Excelencia en Psicología Económica y del Consumo (CEPEC), Temuco, Chile - [ruben.sanchez@ufrontera.cl](mailto:ruben.sanchez@ufrontera.cl)

**Carlos del Valle** [CV] [ [ORCID](http://orcid.org/0000-0002-9905-672X)] <http://orcid.org/0000-0002-9905-672X> [ [GS](#)]

[https://scholar.google.com/citations?user=tgRD\\_30AAAAJ&hl=es&oi=ao](https://scholar.google.com/citations?user=tgRD_30AAAAJ&hl=es&oi=ao)

Universidad de La Frontera, Núcleo Científico Tecnológico en Ciencias Sociales y Humanidades, Doctorado en Comunicación, Temuco, Chile - [carlos.delvalle@ufrontera.cl](mailto:carlos.delvalle@ufrontera.cl)

**Marta Mensa** [CV] [ [ORCID](https://orcid.org/0000-0002-8239-3674)] <https://orcid.org/0000-0002-8239-3674> [ [GS](#)]

<https://scholar.google.com/citations?user=Q2GEVnkAAAAJ&hl=es&oi=ao>

Instituto de Comunicación Social de la Universidad Austral de Chile, Núcleo de investigación sobre Estudios interdisciplinarios en torno a Desigualdad y Derechos Humanos, Valdivia, Chile - [marta.mensa@uach.cl](mailto:marta.mensa@uach.cl)

## Abstracts

[ES] **Introducción.** Los contenidos sobre materia alimentaria difundidos por los medios masivos han ido en aumento en las dos últimas décadas. Dichos contenidos son poderosos elementos estructuradores de la sociedad contemporánea occidental. Este artículo presenta un método para obtener corpus de miles de noticias publicadas en prensa digital –sin tener que recurrir a bases de datos de pago—que puedan ser procesados con software de análisis cuantitativo. **Metodología.** Se operacionaliza el hecho alimentario en una serie de palabras clave que permiten recuperar las noticias sobre el tema en cuestión usando Google y el servicio Import.io. **Resultados.** Se recuperaron en total más de 2500 noticias publicadas en 2016 por tres diarios chilenos y se almacenaron de tal manera que se pueden analizar con técnicas Big Data. **Discusión y conclusiones.** Se comentan las diferencias y

similitudes de los resultados obtenidos para cada diario desde el punto de vista teórico y metodológico. Se mencionan las distintas posibilidades de investigación que ofrecen los datos obtenidos.

**[EN] Introduction.** Food-related news disseminated by the mass media have increased over the last two decades. These contents are powerful structuring elements of Western contemporary society. This article presents a method for the selection of a corpus of thousands of news published in the digital press –without having to resort to paid news databases– that can be subsequently processed with quantitative analysis software. **Methods.** Food-related content is operationalised through a series of keywords that allow for the retrieval of news on the subject in question using the Google search engine and the Import.io service. **Results.** More than 2500 news items published in 2016 by three Chilean newspapers were retrieved and stored for their subsequent analysis with big data techniques. **Discussion and conclusions.** The differences and similarities of the results obtained for each newspaper are discussed from different theoretical and methodological points of view. The different research possibilities offered by the data obtained are also mentioned.

### Keywords

**[ES]** metodología; Big Data; prensa digital; alimentación y periodismo.

**[EN]** Methodology; Big Data; digital press; food and journalism.

### Contents

**[ES]** 1. Introducción. 1.2 Comunicación y alimentación. 1.3 Marco teórico. 1.4 Objetivos. 2 Metodología. 2.1 Operacionalización del hecho alimentario. 2.2 Procedimiento para obtener las noticias. 3. Resultados. 4. Discusión. 5. Conclusión.

**[EN]** 1. Introduction. 1.2. Communication and food. 1.3. Theoretical framework. 1.4. Objectives. 2. Methods. 2.1. Operationalisation of food. 2.2. Procedure to retrieve news. 3. Results. 4. Discussion. 5. Conclusion. 6. Notes. 7. References.

Traducción de **CA Martínez-Arcos**  
(Dr. en Comunicación, Universidad de Londres)

## 1. Introducción

Debido al valor negativo que la filosofía occidental en general, desde Platón a Karl Marx, ha otorgado a la comida y a los placeres fisiológicos (Heldke, 2006; Korthals & Kooymans, 2004; Telfer, 2002), resulta fácil no caer en la cuenta de la verdadera relevancia social, cultural y política que tienen la comida y los discursos –privados o públicos– que se construyen a su alrededor. Afortunadamente y gracias a la antropología y a la sociología, desde principios del siglo XX se viene desarrollando un trabajo académico sobre las dimensiones psicosociales y políticas de la alimentación que ha descubierto en la comida un hecho tan complejo y primordial para la vida humana que desde los albores del siglo XXI se está configurando un campo de estudio interdisciplinar que en los países anglosajones se ha bautizado como “Food Studies”(Poulain, 2017).

El contenido sobre materia alimentaria presente en los medios de comunicación de masas ha ido en aumento en las dos últimas décadas (Contreras & Gracia Arnaiz, 2005, p. 16; Koldobsky, Daniela, 2011, p. 15; Narváez, 2013). Estos discursos son poderosos elementos estructuradores de las sociedades modernas no solo por tratar de un fenómeno humano/social total, sino porque son difundidos por los medios masivos, agentes primordiales en la configuración de la “esfera pública” (Habermas, Domènech, & Grasa, 1981). A pesar de su importancia sociopolítica, el estudio de los contenidos alimentarios presentes en los medios masivos es todavía escaso. Los editores de *Food as*

*Communication / Communication as Food*, un libro publicado en 2011 que recoge las contribuciones de más de 20 académicos sobre alimentación y comunicación, reconocen explícitamente que, si bien el hecho alimentario ha sido analizado ampliamente desde una perspectiva antropológica, sociológica e histórico-cultural, las Ciencias de la Comunicación apenas se han fijado en él (Cramer, Greene, & Walters, 2011).

Con el objetivo de contribuir al estudio del hecho alimentario desde las ciencias de la comunicación, en este artículo ofrecemos herramientas teóricas y metodológicas para desarrollar análisis de contenido en prensa online que permitan, por lo menos, explorar, describir y analizar, la cobertura del hecho alimentario, entendido este último de manera amplia e integral como veremos más adelante. Los autores coinciden con Riffé *et al.* en que un análisis de contenido es un paso previo e imprescindible para estudiar la relación entre dicha cobertura y 1) su producción por parte de los medios; y 2) los efectos que dicha cobertura pudiera tener en las audiencias (2005). Y aún se puede considerar una tercera relación: la que se establece entre los contenidos alimentarios difundidos por los medios masivos y las culturas, ideologías, prácticas..., alimentarias de una determinada sociedad.

Así, el lector encontrará un breve estado del arte sobre alimentación y comunicación, un marco teórico sobre el hecho alimentario a partir de las ciencias sociales, una descripción detallada del procedimiento a seguir para conformar grandes corpus de noticias sobre el hecho alimentario –o de cualquier tema que se estime oportuno– que pueda ser analizado con software, los resultados de la aplicación de dicho método en tres diarios online chilenos, y una discusión sobre los resultados y las posibilidades de investigación que se abren una vez conformado el corpus deseado.

## 1.2 Comunicación y alimentación

La investigación académica de los contenidos no publicitarios sobre materia alimentaria difundidos por los medios de comunicación es todavía incipiente. Por lo general, toma la alimentación como subsidiaria de dos temas que sí reciben atención regular por parte de la comunidad académica: la salud, por ejemplo alimentación, obesidad y trastornos alimentarios (Evans, Rich, Davies & Allwood, 2008; Lawrence, 2004; Menezes Ferreira, de Castro Oliveira & Terrón Blanco, 2015; Plaza, 2012; Sandberg, 2007; Westall, 2011), y el cambio climático, por ejemplo alimentación y calentamiento global (Roig, 2013). Los pocos trabajos sobre medios que se han encontrado que toman la alimentación como único objeto de estudio tienen que ver principalmente con la gastronomía (Blanco Hernández, 2015; Martínez & Poyatos, 2015; Sánchez Gómez, 2010), si bien se ha encontrado un estudio sobre divulgación alimentaria desde el punto de vista de la nutrición (Bernabeu-Peiró, 2015), y otro que analiza cuantitativamente las noticias de alimentación en la prensa madrileña (Fúster, Ribes, Bardón & Marino, 2009).

El semiólogo francés Roland Barthes fue uno de los primeros en estudiar los discursos alimentarios en los medios de comunicación y, especialmente, en la publicidad (Contreras & Gracia Arnaiz, 2005, p. 90). En esta última Barthes (1997) detectó tres motivos principales a la hora de publicitar los alimentos: 1) la rememoración de un tiempo pasado y un espacio rural que reinscriben al consumidor en la historia de su país; 2) la salud por la nutrición (nutrición funcional); y 3) la asociación de alimentos y prácticas alimentarias a situaciones concretas de la vida moderna, por ejemplo, un café para relajarse o un snack para significar a la persona que tiene mucho trabajo y poco tiempo para comer.

Una mirada más amplia y actual a los discursos alimentarios que aparecen en los *mass media* permite distinguir por lo menos cuatro tipos de discursos (Koldobsky, Daniela, 2011): 1) el didáctico, empleado en programas de cocina; 2) el publicitario; 3) el científico-médico, centrado especialmente en el impacto nutricional en la salud de las personas; y 4) la crítica gastronómica. Nosotros añadiríamos un

quinto que por obvio podría pasar desapercibido: el periodístico en el sentido más estricto. Consideramos discursos alimentarios de tipo periodístico a aquellos que informan y analizan sobre su producción, distribución, consumo y eliminación de residuos, considerando especialmente los puntos de vista político-económico y medioambiental.

Una revisión histórica del periodismo alimentario (Albala, 2013) muestra que, por lo menos en el mundo anglosajón, hasta 1957 se escribía mucho sobre la producción y distribución de la comida pero poco o nada sobre los placeres del comer. La crítica gastronómica y la educación culinaria surgieron con fuerza en el último tercio del siglo XX. Finalmente, con la aparición de internet se produjeron dos cambios importantes: 1) la alimentación se convirtió en un tópico de alto interés. Nacieron blogs y webs dedicadas a ello para responder a la gran demanda de la audiencia; y 2) la crisis del papel en la prensa escrita causada por internet hizo que poco a poco el contenido periodístico esté cada vez más en la red. Por consiguiente, el periodismo alimentario, también “se traslada” a internet.

En internet encontramos básicamente tres tipos de publicaciones periódicas –por lo menos en inglés– que cubren la materia alimentaria (Albala, 2013): 1) versiones online de prensa generalista. Estos sitios en general reproducen el mismo contenido que en sus versiones en papel, si bien cada vez más albergan contenidos extra solo disponibles en su versión online. Las secciones dedicadas a la alimentación pueden verse ampliadas en la versión web. 2) Publicaciones especializadas en alimentación solo presentes en la red; y 3) blogs personales sobre materia alimentaria, los cuales, en contadas ocasiones, han cosechado altos índices de audiencia que han permitido a sus editores dedicarse a los mismos en exclusiva y de manera profesional gracias a la publicidad online.

### 1.3 Marco teórico

Los autores consideran que la escasa bibliografía encontrada sobre los discursos alimentarios en los *mass media* no se justifica si tomamos en consideración que en la actualidad, dichos discursos no solo están ampliando el conocimiento del ciudadano de a pie sobre el complejo fenómeno alimentario sino que están reorganizando nuestra relación con la tierra y con nuestros cuerpos (Frye & Bruner, 2013). Desde la semiótica concretamente, autores como Roland Barthes (1997), Jean Soler (1997), Gastón Gaínza (2003) y Fabio Parsecoli (2011) han mostrado que las prácticas alimentarias junto con los alimentos que se degustan, son un sistema de comunicación que, al igual que el lenguaje verbal, refleja un mundo sociocultural determinado. La comida y el comer, por lo tanto, son signos; son textos ubicados en la semiosfera particular de cada comunidad y que juegan un papel fundamental en la reproducción tanto biológica como cultural de la sociedad. Visto desde la sociosemiótica, los textos alimentarios comunican ideologías que configuran relaciones de poder.

Tan importante como la capacidad de simbolizar que tienen los alimentos son los discursos que construimos alrededor del hecho alimentario. La razón es obvia: si los alimentos “hablan” de la sociedad que los produce, los prepara y los degusta, cuando la sociedad habla de los alimentos, al fin y al cabo, está hablando de sí misma. Así, los discursos sobre la comida son otro mecanismo de socialización por el que entendemos y construimos nuestra cultura (Cramer *et al.*, 2011). Por lo tanto, también podemos referirnos a las formas de hablar y de representar el hecho alimentario como poderosos significantes ideológicos que construyen relaciones de poder (LeBesco & Naccarato, 2008). Así lo cree Thompson (2012) quien llega a equiparar los emergentes discursos sobre materia alimentaria a los discursos sobre Derechos Humanos que surgieron en la segunda mitad del siglo XX. Y es que según Thompson, el discurso sobre el alimento se está convirtiendo en un lenguaje de lucha política y cultural que trasciende la materia alimentaria. Es en los discursos sobre la comida donde se negocia el poder, se buscan identidades y capacidad de acción (“agency”) en el mundo globalizado actual.

Los contenidos alimentarios difundidos por los medios de comunicación de masas son especialmente relevantes porque estos últimos son una institución paradigmática y cultural de las sociedades modernas (J. B. Thompson, 1998) que, mediante los contenidos que producen y difunden, determinan los límites y horizontes de los mundos posibles (Alsina, 1989; Farré, 2004). Dichos contenidos estructuran y promueven significaciones socioimaginarias que orientan, ordenan, clasifican y organizan los hechos sociales en los que se involucran los sujetos (Dijk, 2009). Por ello, la Antropología de la Comunicación dirá que los *mass media* son la “estructura de acogida” –agente educador y enculturador– moderna, la cual hoy en día viene a complementar, y hasta a suplantar, las tres estructuras de acogida tradicionales: la familia, la ciudad y la iglesia (Duch & Chillón, 2012). El poder, pues, que tienen los medios de comunicación como instituciones productoras de formas simbólicas para modelar y transformar la sociedad es evidente (Velázquez, 1992). Tanto que, a decir de Thompson (1998) y Bourdieu (2001), son los *mass media*, y no otro, el agente, la institución, que posee y administra el poder de tipo simbólico que circula por los media en las sociedades mediatizadas como la actual.

Para estudiar los contenidos alimentarios difundidos por los *mass media* proponemos conceptualizar el hecho alimentario de una manera holística siguiendo los aportes teóricos generados principalmente por la antropología y la sociología. En la actualidad se considera el “hecho alimentario” (Contreras & Gracia Arnaiz, 2005) un fenómeno social total, en el sentido maussiano del término, y un fenómeno humano total, en palabras del filósofo y sociólogo francés E. Morin (2014). “Social total” porque en la alimentación encuentran expresión simultánea todas las áreas de la cultura y tipos de instituciones (económicas, jurídicas, políticas, religiosas, etc.) (Elias, 1989; Goody & Willson, 1995); y “humano total” por sus componentes biológicos y ecológicos, y por su poder estructurante de la “organización social, situándose en el mismo nivel de importancia, o incluso mayor, que la sexualidad o el parentesco” (Contreras & Gracia Arnáiz, 2005). Sobre esto último destacan especialmente los trabajos de Claude Lévi-Strauss (1981, 2013; 1970), Mary Douglas (1972, 1980) y Pierre Bourdieu (1984). Este último, en su reconocido trabajo sobre la distinción social, llegó a postular que los hábitos alimentarios son una manera de naturalizar no solo la diferenciación social, sino que también la ideología.

Esta concepción amplia del hecho alimentario permite abordar los contenidos alimentarios difundidos por los medios masivos más allá de la nutrición y la gastronomía. Entender la alimentación como “hecho social total”, también nos permite conjugar dos conceptos clave sobre la alimentación humana: “sistema alimentario” y “cultura alimentaria”. De esta forma, se puede trabajar con una definición integral de la alimentación que incluya tanto su dimensión simbólica como material. Así, Por “sistema alimentario” hay que entender como mínimo [1] el conjunto de relaciones que se establecen entre la producción, distribución, preparación, consumo y eliminación de los alimentos (Goody & Willson, 1995). Y, por “cultura alimentaria”, “el conjunto de representaciones, de creencias, conocimientos y de prácticas heredadas y/o aprendidas que están asociadas a la alimentación y que son compartidas por los individuos de una cultura dada o de un grupo social determinado dentro de una cultura.” (Contreras & Gracia Arnaiz, 2005) Solo así, considerando las múltiples dimensiones del hecho alimentario, se podrá caracterizar cuantitativa y cualitativamente, y de manera integral, la cobertura mediática en materia alimentaria difundida por los *mass media*.

#### 1.4 objetivos

El objetivo de este artículo es, entonces, explicar cómo conformar un corpus suficientemente amplio y representativo a partir de prensa online que permita llevar a cabo, por lo menos, estudios exploratorios-descriptivos sobre la cobertura del hecho alimentario utilizando como método el análisis de contenido y apoyándose en técnicas de *Big Data*. Por análisis de contenido entendemos lo que

Bernard Berelson (1952) definió como una técnica de investigación para la descripción objetiva, sistemática y cuantitativa del contenido manifiesto de la comunicación. Por contenido manifiesto o denotativo se entiende el significado del símbolo en cuestión en el que tanto emisor como receptor estarían espontáneamente de acuerdo. Una forma más actual de definir el análisis de contenido es la que proveen Riffe *et al.* (2005) para quien éste consiste en “la asignación sistemática de categorías a contenido comunicacional según unas reglas preestablecidas, y así poder analizar las relaciones entre categorías usando métodos estadísticos”. Esto con el objetivo de realizar “inferencias reproducibles y válidas que puedan aplicarse a su contexto” (Krippendorff, 1990).

## 2. Metodología

El aumento de contenidos sobre alimentación que Internet ha permitido, junto con las posibilidades informáticas desarrolladas para analizar grandes cantidades de datos (*Big Data*), invita a considerar corpus de cientos o miles de noticias (unidades de análisis) con el objetivo de obtener una caracterización general de cobertura sobre el hecho alimentario que los medios hacen en general, más allá de eventos, vicisitudes o crisis puntuales que se puedan dar. Este tipo de análisis permitirá observar de *qué* se habla mucho, poco y nada (por ejemplo, *agenda setting*), antes de centrarse en análisis discursivos más interesados en el *cómo* se habla del sistema y la cultura alimentaria (por ejemplo, teoría de *framing*).

La generación de un corpus de cientos o miles de noticias (o entradas en un blog) publicadas en formato digital se puede conseguir con relativa facilidad usando el buscador Google, si bien el proceso, como veremos, es un tanto laborioso. Es verdad que, para algunos medios de comunicación, existe la posibilidad de obtener las noticias que cumplan con los criterios previamente establecidos a través de los agregadores de contenidos de prensa como LexisNexis y Factiva. Sin embargo, estos agregadores presentan dos desventajas: 1) son de pago y, por tanto, no todas las universidades o centros de investigación proveen este servicio a sus respectivos investigadores; y 2) no todos los medios informativos, especialmente latinoamericanos, están presentes en dichos agregadores.

Al entender de los autores, la mejor alternativa a estos agregadores es Google porque es un servicio gratuito y, por ende, disponible para toda la comunidad académica; y, lo más importante, porque se ha documentado su fiabilidad a la hora de buscar noticias –la literatura científica ha establecido que las búsquedas de noticias en lengua inglesa en Google y en LexisNexis presentan un grado de coincidencia superior al 80% (Weaver & Bimber, 2008).

### 2.1 Operacionalización del hecho alimentario

Ya sea que el investigador use Google o una base de datos como LexisNexis con el objetivo de conformar el corpus de investigación, es necesario operacionalizar la categoría “hecho alimentario” mencionada en la introducción en una serie de palabras clave. Para ello se ha partido de un análisis de la discusión teórica que los antropólogos de la alimentación Jesús Contreras y Mabel Gracia realizan sobre el sistema alimentario en *Alimentación y Cultura* (2005), y de dos investigaciones cuantitativas realizadas en España sobre contenidos alimentarios en medios de comunicación (Fúster *et al.*, 2009; Marín-Murillo, Flora, Armentia-Vizueté, José-Ignacio & Olabarri-Fernández, Elena, 2016). Esta forma de obtener las palabras clave que permitirá obtener las noticias relevantes para conformar el corpus deseado, está en consonancia con las buenas prácticas recomendadas para el análisis de contenido (Lacy, Watson, Riffe & Lovejoy, 2015).

Tomando los cinco procesos fundamentales del sistema alimentario descritos por Goody (1995), producción, distribución, preparación, consumo y limpieza/eliminación de residuos, se ha generado una matriz con cinco columnas, en la que cada columna corresponde a un proceso. Luego, con el apoyo

de la discusión teórica de Contreras et al. (2005), se han añadido subcategorías a algunos procesos con el objetivo de reconocer la complejidad de los mismos. Finalmente, con la ayuda de a) investigaciones previas sobre contenidos alimentarios en los medios de comunicación, b) de literatura científica sobre antropología y sociología de la alimentación, c) de la experiencia anecdótica provista por la lectura de contenidos alimentarios en los tres diarios digitales chilenos seleccionados que más adelante identificaremos, y d) de la triangulación entre el investigador y coinvestigadores de este proyecto, se han identificado los términos clave que mejor denotan los procesos ya mencionados del sistema alimentario.

Una vez identificadas las palabras clave que operacionalizan el “hecho alimentario”, se han eliminado aquellas palabras que se repiten en más de un proceso. De esta manera se simplifica la búsqueda sin arriesgar la omisión de contenidos relevantes. Por ejemplo, en lugar de solicitar a Google que nos devuelva noticias que contengan las expresiones “alimentación” y “alimentación saludable”, simplemente se solicita que busque por “alimentación”, ya que una búsqueda únicamente motivada por esta última palabra también devolverá las noticias que contengan la expresión “alimentación saludable”.

Así, el listado de términos que se proponen para obtener las noticias relevantes para el corpus deseado son los siguientes:

Alimento/s; alimentación; alimentario/a; biocombustible; agroindustria; transgénicos; semilla; “explotaciones agrícolas”; pesca; pescado; piscicultura; acuicultura; pesquera; frutícola; vinícola; ganadería; agricultura; supermercado; cocina; cocinar; culinario/a; dieta; nutrición; nutrientes; malnutrición; nutricional; comida; comer; hambre; bebida/s; beber; gastronomía; gastronómico/a; degustación; restaurante; cocinería; picá; chef; pastelería; repostería.

Los términos y expresiones relativos a la basura y residuos alimentarios no se incluyen en el listado de términos de búsqueda porque, primero, “basura” es un término demasiado genérico que provocaría que Google nos ofreciera un listado de resultados lleno de contenidos irrelevantes para nuestra investigación; y segundo, porque a partir de nuestra experiencia, siempre que se habla de residuos alimentarios, aparecen palabras como “comida” y “alimentos”, términos que sí están en nuestro listado de búsqueda.

## 2.2 Procedimiento para obtener las noticias

Con el objetivo de ilustrar el procedimiento para obtener las noticias usando Google, a continuación se detallan los pasos seguidos para conseguir las noticias publicadas en 2016 por el diario online chileno más visitado [2]: [www.emol.com](http://www.emol.com). Este diario sería una unidad de contexto (Colle, Raymond, 2011) de donde obtener las unidades de análisis que terminarán conformando el universo de noticias a analizar.

Si bien algunos diarios online tienen buscadores propios, Google es la mejor opción por dos razones: 1) No es posible saber la fiabilidad de un buscador de diario digital; y 2) en caso de querer comparar los contenidos de distintos diarios online, las más que probables diferencias en fiabilidad y funcionamiento de cada buscador particular respectivo, comprometen la confiabilidad de los datos para analizar. Google, en cambio, puede actuar de árbitro entre diarios online a la hora de saber qué contenidos ha publicado cada medio escrito.

Para que Google arroje resultados despersonalizados –es decir, que no estén condicionados por la huella digital en internet del investigador– y lo más ajustados a la realidad del diario online en cuestión, es necesario seguir los siguientes pasos. En primer lugar, hay que crear una nueva cuenta de usuario en el sistema operativo del computador que vaya a emplearse para este fin. Una vez creada, es necesario reiniciar el computador e ingresar con esta nueva cuenta. A continuación, se ejecuta el navegador de preferencia (recomendamos Firefox), y abriremos una ventana de navegación privada (en Firefox, “Nueva ventana privada” y en Chrome “Nueva ventana de incógnito”).

En la barra de búsqueda, escribir “buscador Google” y el país en que estén radicados los medios online a estudiar. En nuestro caso, escribimos “buscador Google Chile”. Una vez se accede al buscador Google del país correspondiente, es necesario configurarlo de manera correcta para el trabajo de búsqueda. Para ello, se accede al menú “Preferencias” situado en la parte baja y derecha de la pantalla. En “Preferencias”, el investigador debe asegurarse de que la modalidad “Activar SafeSearch” no esté seleccionada. Luego, en “Resultados por página”, elegir “100”; luego pinchar en “historial de búsqueda” y asegurarse de que esté desactivado; a continuación, seleccionar la región en la que está radicado el diario. Esto es imprescindible para forzar a Google a que nos arroje resultados almacenados en los servidores del país en el que está radicado el diario en cuestión. Debido a que Google ofrece resultados según la localización del computador que genera la búsqueda, haciendo esto uno se asegura de obtener los resultados más parecidos a los que obtendrían los ciudadanos del país del diario online que pretendemos estudiar.

Una vez realizada esta configuración previa, es necesario acceder a “búsqueda avanzada”. Para ello se vuelve al menú “Preferencias” y se pincha en “Búsqueda avanzada”. Este servicio permite el uso de booleanos, además de restringir la búsqueda a un único dominio web, entre otras opciones. En el campo “todas estas palabras” hay que escribir, para empezar, la primera palabra clave de nuestro listado. En nuestro caso, “alimento”. En el campo “sitio o dominio”, se introduce el dominio del medio de comunicación que se quiera estudiar (en nuestro caso, “[www.emol.com](http://www.emol.com)”). A continuación, se presiona el botón azul que reza “Búsqueda avanzada” y Google devolverá una página con los primeros 100 resultados. En la parte superior de los resultados e inmediatamente debajo de la barra de búsqueda se encuentra un menú con varias opciones. Al presionar en “Herramientas” aparece un nuevo menú que permite filtrar los resultados para que solo se muestren las noticias publicadas entre el 1 de enero de 2016 y el 31 de diciembre del mismo año.

Si el investigador ha seguido los pasos hasta aquí descritos, tendría que encontrarse en la página de resultados de Google con la siguiente expresión en la barra de búsqueda: “alimento site:www.emol.com”. A continuación, viene el paso más importante para asegurar que Google arroja resultados despersonalizados. Según especialistas en márketing (Nedelko, 2013), este paso es tan efectivo que algunos de los anteriores serían innecesarios. Consiste en lo siguiente. Una vez se tiene la primera página con los 100 primeros resultados –habrá palabras clave que arrojarán menos de 100 resultados—en la barra del navegador (no confundir con la barra de búsqueda de Google) hay que pegar al final de la URL la siguiente expresión sin las comillas: “&pws=0” [3] y presionar la tecla “Enter”. Esta expresión dice obliga a Google a que entregue resultados despersonalizados. A modo de ejemplo, se presentan a continuación las dos URLs de la misma búsqueda, la primera sin la expresión para despersonalizar, y la segunda con dicha expresión resaltada en negrita:

Es importante notar que esta búsqueda con la expresión que pide a Google resultados despersonalizados, solo es válida para la primera página de resultados. Cuando se acceda a la segunda página de resultados de una misma palabra clave, hay que volver a pegar la expresión “&pws=0” en la barra del navegador.



A partir de ahora, tan solo cambiando el término de búsqueda, Google arrojará las noticias encontradas de 100 en 100, publicadas entre las fechas señaladas en el dominio [www.emol.com](http://www.emol.com). Por ende, no sería necesario repetir todos los pasos de configuración recién explicados por cada una de las palabras clave anteriormente listadas. Sí que habría que añadir en todos los casos la expresión “&pws=0” para asegurarse de recibir resultados despersonalizados.

URLs para alimento site:www.emol.com entre 1-1-2016 y 31-12-2016
--

<a href="https://www.google.es/search?q=alimento+site%3Awww.emol.com&amp;num=100&amp;lr=&amp;hl=es-419&amp;as_qdr=all&amp;source=Int&amp;tbs=cdr%3A1%2Ccd_min%3A1%2F1%2F2016%2Ccd_max%3A12%2F31%2F2016&amp;tbm=">https://www.google.es/search?q=alimento+site%3Awww.emol.com&amp;num=100&amp;lr=&amp;hl=es-419&amp;as_qdr=all&amp;source=Int&amp;tbs=cdr%3A1%2Ccd_min%3A1%2F1%2F2016%2Ccd_max%3A12%2F31%2F2016&amp;tbm=</a>
---

<a href="https://www.google.es/search?q=alimento+site%3Awww.emol.com&amp;num=100&amp;lr=&amp;hl=es-419&amp;as_qdr=all&amp;source=Int&amp;tbs=cdr%3A1%2Ccd_min%3A1%2F1%2F2016%2Ccd_max%3A12%2F31%2F2016&amp;tbm=&amp;pws=0">https://www.google.es/search?q=alimento+site%3Awww.emol.com&amp;num=100&amp;lr=&amp;hl=es-419&amp;as_qdr=all&amp;source=Int&amp;tbs=cdr%3A1%2Ccd_min%3A1%2F1%2F2016%2Ccd_max%3A12%2F31%2F2016&amp;tbm=&amp;pws=0</a>
---

Idealmente, debería realizarse una búsqueda por cada palabra clave. Sin embargo, con el objetivo de aligerar el proceso de búsqueda, es posible usar más de una palabra clave en cada búsqueda. De hacerse así, hay que tener en cuenta las siguientes consideraciones. Primero, asegurarse de que usamos el boleano OR y no AND –este último se usa por defecto en google cuando escribimos varias palabras en la barra de búsqueda. Para ello, en la página de búsqueda avanzada, en lugar de llenar el campo “todas estas palabras”, hay que escribir separadas por coma las palabras clave que se desee buscar en el campo “cualquiera de estas palabras”. Otra manera de hacerlo es directamente en la barra de búsqueda de la página de resultados, separando cada palabra con una coma y el boleano OR, por ejemplo, “alimento, OR alimentación”. La segunda consideración que debe tenerse muy en cuenta es que Google ofrece como máximo 500 resultados. Es decir, que si por ejemplo, la búsqueda de la palabra clave “alimentación” entre las noticias publicadas por [www.emol.com](http://www.emol.com) durante 2016 arrojara como resultado más de 500 noticias que contienen dicha palabra, Google solo mostraría 500. Es por ello que, si se opta por buscar más de una palabra a la vez, es necesario asegurarse de que Google devuelve menos de 500 resultados, porque si no, podríamos estar perdiendo noticias. Si la búsqueda de una palabra clave diera más de 500 resultados, sería necesario dividir en varias búsquedas el espacio temporal considerado. Por ejemplo, en lugar de pedir a Google que arroje resultados de 2016, se le solicita en una primera búsqueda que considere solo los seis primeros meses de dicho año. Finalmente, hay que señalar que, ya sea que se realice una nueva búsqueda por cada palabra clave, o búsquedas con más de una palabra clave, al juntar los resultados de todas las búsquedas se obtendrá un número importante de noticias duplicadas. Como se verá más adelante, la obtención de resultados duplicados es un problema menor que fácilmente se resuelve en una fase más avanzada del proceso de conformación de corpus.

Una vez se tenga la página de resultados de Google con 100 noticias publicadas en [www.emol.com](http://www.emol.com) durante 2016, hay dos opciones para obtener cada una de las noticias arrojadas por Google. O bien se pincha en cada una de ellas y se procede a copiar los contenidos en una hoja de Excel, o bien trabaja con herramientas informáticas que puedan automatizar este proceso. La segunda opción es mucho más ágil y fiable, y por lo tanto, hay que optar por ella.

La opción de usar herramientas informáticas para obtener las noticias puede materializarse de distintas formas. Aquí se propone una manera que requiere de un servicio online de pago de bajo costo que permite obtener las URLs de las noticias arrojadas por Google, y luego también el contenido de dichas URLs, pues es su contenido el que interesa analizar.

El servicio web que los autores emplearon y recomiendan para obtener las URLs de <http://www.emol.com> y de otros dos diarios online chilenos es <http://www.import.io>. Si bien este servicio es de pago, ofrece la posibilidad de usarlo por un tiempo limitado y un número limitado de “trabajos” sin coste alguno.[4] Su principal desventaja es que solo está en inglés. La forma de trabajar con este servicio es la siguiente: el usuario entrega X número de URLs al sistema y le pide que extraiga los elementos deseados de cada una de las webs a las que llevan las URLs entregadas. En el caso que nos ocupa, se entregó a [www.import.io](http://www.import.io) las URLs de cada una de las páginas de resultados ofrecidas por Google por cada búsqueda. Es decir, si una búsqueda determinada en Google con X palabras clave dio como resultado entre 200 y 300 noticias, se recogieron las tres URLs que pueden encontrarse en la barra del navegador—no la barra de búsqueda de Google—de las tres páginas de resultados respectivas. Una vez entregadas dichas URLs, se instruyó a [www.import.io](http://www.import.io) para que recogiera el título, la URL, la fecha de publicación y la descripción de cada resultado y organizara la información recogida en filas y columnas, donde cada fila es un resultado o caso, y cada columna un aspecto del resultado, es decir, una columna para título, otra para la URL, otra para fecha y otra para la descripción.

Una vez terminado este proceso para todas las búsquedas realizadas en Google, se juntan todas las hojas de cálculo producidas por [import.io](http://import.io) en un único archivo usando software adecuado como Microsoft Excel o Libreoffice. A continuación, será necesario depurar los resultados mediante un doble procedimiento. Primero, se seleccionará la columna de URLs y se solicitará al software que elimine los duplicados. Segundo, se ordenen los resultados por fecha para comprobar que efectivamente solo hay noticias publicadas en 2016. El tercer paso requiere de la intervención humana porque consiste en detectar falsos positivos, es decir, en identificar aquellas URLs que efectivamente contienen alguna de las palabras clave que se han buscado, pero que, en realidad, la noticia que contiene la URL en cuestión, o no tiene nada que ver con el hecho alimentario, o simplemente el hecho alimentario no es ni tema principal ni secundario. A continuación, se presentan tres ejemplos del [www.emol.com](http://www.emol.com) para ilustrar los falsos positivos:

Ejemplo 1:

Título del resultado según Google:

“San Fermín 2016: Tercera jornada deja a 13 heridos, dos de ellos por ...”

URL del resultado:

<http://www.emol.com/noticias/Internacional/2016/07/10/811755/San-Fermin-2016-Tercera-jornada-deja-a-13-heridos-dos-de-ellos-por-asta-de-toro.html>

Fecha de la noticia según Google:

Jul 10, 2016 -

Descripción del resultado según Google:

“Jul 10, 2016 - Para este domingo, a las 02 AM hora local, participa la ganadería salmantina de Pedraza de Yeltes, quienes debutan en esta versión de San ...”

Este es un ejemplo de falso positivo porque si bien la noticia contiene palabras clave, el tema son las fiestas taurinas que se celebran en España, y no la alimentación humana.

Ejemplo 2

Título del resultado según Google:

“CONCURSO: ¿Te gusta la cocina "simple"? Tenemos ganadora del ...”

URL del resultado:

<http://www.emol.com/noticias/Tendencias/2016/05/09/801927/CONCURSO-Te-gusta-la-cocina-simple-Comparte-tu-receta-mas-facil-y-gana-un-robot-Moulinex.html>

Fecha de la noticia según Google:

May 9, 2016 -

Descripción del resultado según Google:

May 9, 2016 - Para quienes buscan hacer la vida más fácil, la Multicook de Moulinex se presenta como la alternativa ideal para ahorrar tiempo y cocinar de manera fácil y ...

Este es otro ejemplo de falso positivo porque se trata de una “noticia sobre cocina” que en realidad es publicidad encubierta de un utensilio de cocina mediante un concurso.

Ejemplo 3

Título del resultado según Google:

“Protagonista de la serie "Scandal" es homenajeada en Harvard | Emol ...”

URL del resultado:

<http://www.emol.com/noticias/Espectaculos/2016/01/29/785909/Protagonista-de-la-serie-Scandal-es-homenajeada-en-Harvard.html>

Fecha de la noticia según Google:

Jan 29, 2016 -

Descripción del resultado según Google:

“Jan 29, 2016 - Además, Washington se llevó el premio Hasty Pudding, un trofeo en forma de olla dorada para cocinar el pudín. El grupo de teatro, el más antiguo del país, ...”

Este es también un ejemplo de falso positivo porque si bien la noticia contiene la palabra clave “cocinar”, el hecho alimentario no es tema principal ni secundario en esta noticia: es simplemente anecdótico.

Para llevar a cabo la identificación de falsos positivos, se formó a tres codificadores que pudieran evaluar cada noticia según los conceptos de “hecho alimentario”, “sistema alimentario” y “cultura alimentaria” que se han presentado en la introducción de este artículo. Después, se les pidió que juzgaran la pertinencia o no pertinencia al corpus siguiendo los siguientes criterios de exclusión: 1) “falsas noticias”: contenidos como publicidad encubierta, concursos, etc., en los que no hay intención de informar al público; 2) noticias en las que la alimentación no sea tema principal ni secundario. Es decir, noticias en las que el hecho alimentario sea anecdótico; 3) noticias en las que las bebidas alcohólicas sean el motivo principal, a excepción del vino y la cerveza; 4) noticias en formato no escrito como audio, fotografía y vídeo. Y finalmente, con la intención de abarcar el mayor número

posible de noticias, se indicó a los codificadores que ante la duda frente a la pertenencia o no de una noticia al corpus, optaran por considerarla pertinente.

Así, para cada medio de comunicación, se asignaron dos codificadores que evaluaron de manera independiente cada noticia. La evaluación debía hacerse únicamente usando la hoja de cálculo producida por [www.import.io](http://www.import.io) que contenía los elementos de cada caso. Si con dicha información no era posible tomar una decisión, los codificadores cargaban la URL de la noticia en el navegador para leer la noticia completa y tomar una decisión final. Si se consideró pertinente, la fila quedaba en blanco. Si se consideró no pertinente, la fila de la noticia se marcaba en rojo. Y en los casos de duda, se marcaba en amarillo. Una vez finalizado este proceso, un tercer codificador y uno de los autores de este artículo revisaron las noticias señaladas en rojo y amarillo y tomaron una decisión final sobre su pertinencia.

Una vez depuradas todas las URLs para cada diario online, se vuelve a emplear [www.import.io](http://www.import.io) con el objetivo de recoger los elementos de la noticia que se quieran analizar. En nuestro caso, se recogieron los siguientes elementos: diario, fecha, autor, localidad desde donde se escribe la noticia, titular, subtítulo, texto de la noticia, y sección. Así, entregando a [www.import.io](http://www.import.io) las URLs depuradas de cada diario online, y configurando este servicio para que recoja los elementos requeridos de cada noticia, se obtiene un Excel (también se pueden descargar los datos en formato CSV y NDJSON) en el que cada fila es una unidad de análisis (una noticia) y cada columna es una variable distinta de cada unidad de análisis. Hay que señalar que, en el proceso de extracción de noticias a partir de las URLs, algunas URLs pueden “perdersé” por distintos motivos, como por ejemplo, porque alguna URL haya caducado, o porque alguna URL lleve a un documento PDF, o DOC, que obviamente, ya no podría considerarse una unidad de análisis. Especialmente para evitar lo primero, se recomienda que la búsqueda realizada en Google sea de noticias publicadas lo más recientemente posible.

Llegados a este punto, ya se dispondría del corpus o universo de datos, y se podría pasar a la etapa de muestreo o directamente de análisis si se emplea software especializado para el análisis de grandes cantidades de datos textuales. Se puede encontrar más información sobre software para análisis de Big Data textual en la página 304 y ss., de la segunda edición de *The Content analysis* escrito por K. A. Neuendorf (2016). Finalmente, se asegura al lector que el procedimiento aquí desarrollado solo requiere de conocimientos de ofimática básicos. Por lo tanto, está al alcance de todo investigador familiarizado con el uso de la tecnología cotidiana actual.

### 3. Resultados

Con el objetivo de obtener un corpus representativo de los contenidos alimentarios que potencialmente tienen más audiencia, se aplicó el método que acabamos de describir a los siguientes tres diarios online chilenos: <http://www.emol.com>, <http://www.latercera.com>, y <http://www.elmostrador.cl>. El primero es la versión online del diario *El Mercurio*. Fundado en 1827 en la ciudad de Valparaíso, cuenta actualmente con varias ediciones en el país. La edición de la capital, Santiago de Chile, se publicó por primera vez en 1900. *El Mercurio* forma parte del *Grupo El Mercurio*, uno de los dos principales conglomerados comunicativos de Chile. Según los datos de audiencia proporcionados por Alexa y Similarweb en julio de 2016, su versión online es la web de noticias más leída en Chile. La web [www.latercera.com](http://www.latercera.com) es la versión online del diario *La Tercera*, fundado en 1950. Forma parte del *Grupo Copesa*, el otro gran conglomerado comunicativo chileno. Según los datos de Alexa y Similarweb, [www.latercera.com](http://www.latercera.com) era el segundo o tercer medio online más visitado, también a julio de 2016. Finalmente, [www.elmostrador.cl](http://www.elmostrador.cl) es un diario independiente publicado exclusivamente en línea que pertenece a la empresa La Plaza SA. Según los medidores de audiencia ya citados, a julio de 2016 este diario era la séptima web de noticias más consultada.

Los resultados de la aplicación del método que se propone para formar corpus sobre el hecho alimentario a partir de prensa online, fueron los siguientes. La búsqueda en Google arrojó 4318 casos de [www.emol.com](http://www.emol.com), 4462 de [www.latercera.com](http://www.latercera.com), y 5897 de [www.emol.com](http://www.emol.com). La eliminación de URLs duplicadas dejó al [www.emol.com](http://www.emol.com) con 3032 casos, al [www.latercera.com](http://www.latercera.com) con 2739 casos, y al [www.elmostrador.cl](http://www.elmostrador.cl) con 3313 casos. La eliminación de casos por título duplicado (esto significa que una misma noticia tiene dos URLs distintas) dejó al diario [www.emol.com](http://www.emol.com) con 3024 casos, al [www.latercera.com](http://www.latercera.com) con 2715 casos, y al [www.emol.com](http://www.emol.com) con 2949 casos. Después del proceso de evaluación realizado por los revisores, el total de noticias que tienen que ver con el hecho alimentario publicadas en 2016 por diario fueron: [www.emol.com](http://www.emol.com), 961; [www.latercera.com](http://www.latercera.com), 1029; y [www.elmostrador.com](http://www.elmostrador.com), 695. Finalmente, el total de noticias que se consiguió extraer con [www.import.io](http://www.import.io) es el siguiente: [www.emol.com](http://www.emol.com), 912; [www.latercera.com](http://www.latercera.com), 957; [www.elmostrador.cl](http://www.elmostrador.cl), 682. Así, el corpus de esta investigación quedó conformado por un total de 2551 noticias publicadas por los tres medios en cuestión durante 2016. (Ver tabla 1)

<b>Resultados del método propuesto en tres diarios chilenos</b>	
<b>Emol.com</b>	<b>Total</b>
Total noticias arrojadas por Google	4.318
Total noticias sin URLs duplicadas	3.032
Total noticias sin títulos duplicados	3.024
Total noticias sin falsos positivos	961
Total noticias extraídas por <a href="http://www.import.io">www.import.io</a>	912
<b>Latercera.com</b>	
Total noticias arrojadas por Google	4.462
Total noticias sin URLs duplicadas	2.739
Total noticias sin títulos duplicados	2.715
Total noticias sin falsos positivos	973
Total noticias extraídas por <a href="http://www.import.io">www.import.io</a>	957
<b>Elmostrador.cl</b>	
Total noticias arrojadas por Google	5.897
Total noticias sin URLs duplicadas	3.313
Total noticias sin títulos duplicados	2.949
Total noticias sin falsos positivos	695
Total noticias extraídas por <a href="http://www.import.io">www.import.io</a>	682
Fuente: elaboración propia	

El “total de noticias arrojadas por Google”, se refiere al número total de entradas que devolvió Google al conjunto de búsquedas realizadas con el listado de palabras clave presentado en la metodología. Por “total de noticias sin URLs duplicadas”, hay que entender el número total de casos resultado del proceso de detección y eliminación de URLs duplicadas realizado por Excel. Con “total noticias sin títulos duplicados”, se indica el número total de casos resultado del proceso de detección y eliminación de títulos duplicados realizado por Excel. Y el “total noticias sin falsos positivos”, se refiere al total de noticias que se consideran una vez los codificadores han evaluado la pertinencia de cada caso. Finalmente, el “total de noticias extraídas por [www.import.io](http://www.import.io)” señala el número de noticias que se han podido extraer y que conformarán el corpus.

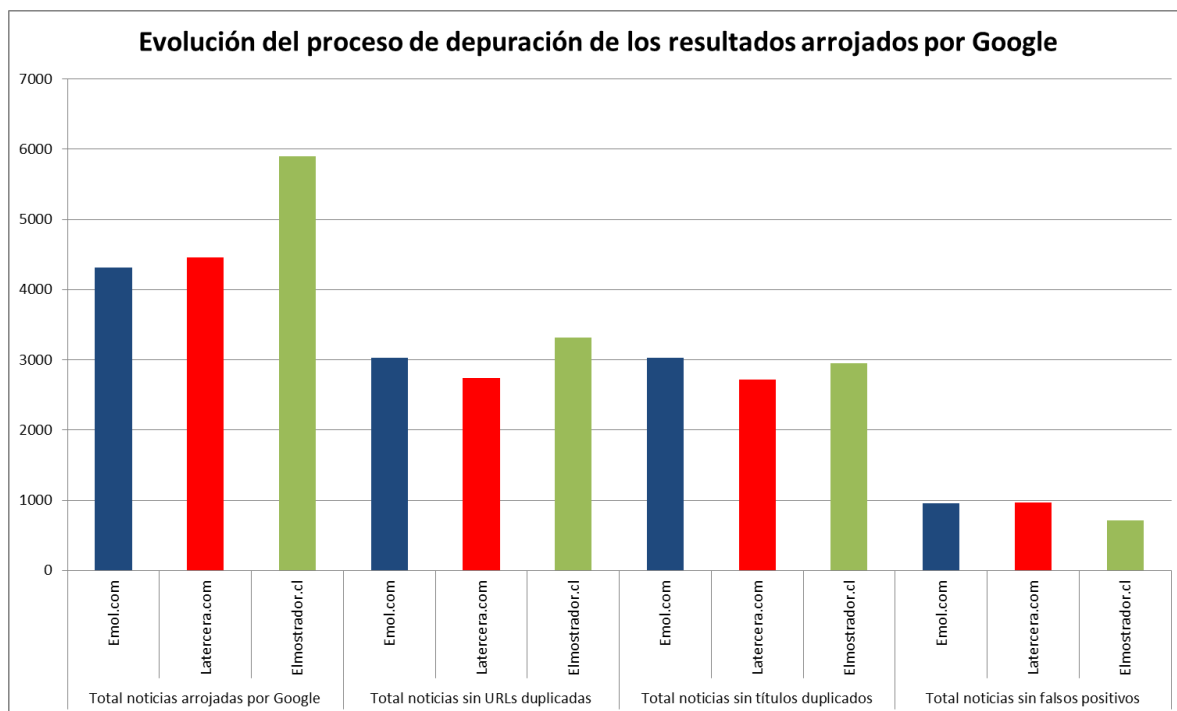


GRÁFICO 1 – Fuente: elaboración propia

En el gráfico 1 se puede observar y comparar cuantitativamente la evolución de este proceso para los tres diarios online que estamos estudiando. En el mismo se observa que el [www.emol.com](http://www.emol.com) y [www.latercera.com](http://www.latercera.com) presentan números muy similares y por ende, una evolución casi idéntica. No así en el caso de [www.elmostrador.cl](http://www.elmostrador.cl). Si bien la búsqueda en Google arrojó para este diario un 33% aprox. de resultados más que el resto de diarios considerados, el total de noticias que finalmente entró en el corpus fue aproximadamente un 30% menos. En la discusión se ofrecen algunas hipótesis de por qué esta diferencia significativa.

Los resultados también nos permiten contabilizar y comparar cuántas noticias sobre alimentación han publicado estos tres diarios durante cada mes de 2016. En el gráfico 2 se puede observar la evolución de los tres diarios. El gráfico muestra tendencias parecidas para los tres diarios, si bien destacan los meses de mayo y octubre por motivos distintos. Si en mayo ambos diarios aumentan de manera significativa el contenido de noticias publicadas sobre el hecho alimentario, ofreciendo además una cantidad de noticias similar, en octubre es cuando se muestra la mayor disparidad cuantitativa, siendo

el [www.elmostrador.cl](http://www.elmostrador.cl) el diario que, siempre según nuestra búsqueda y procedimiento, menos noticias publicó (38) y [www.latercera.com](http://www.latercera.com) el que más (116), dejando al [www.emol.com](http://www.emol.com) en tierra de nadie (67).

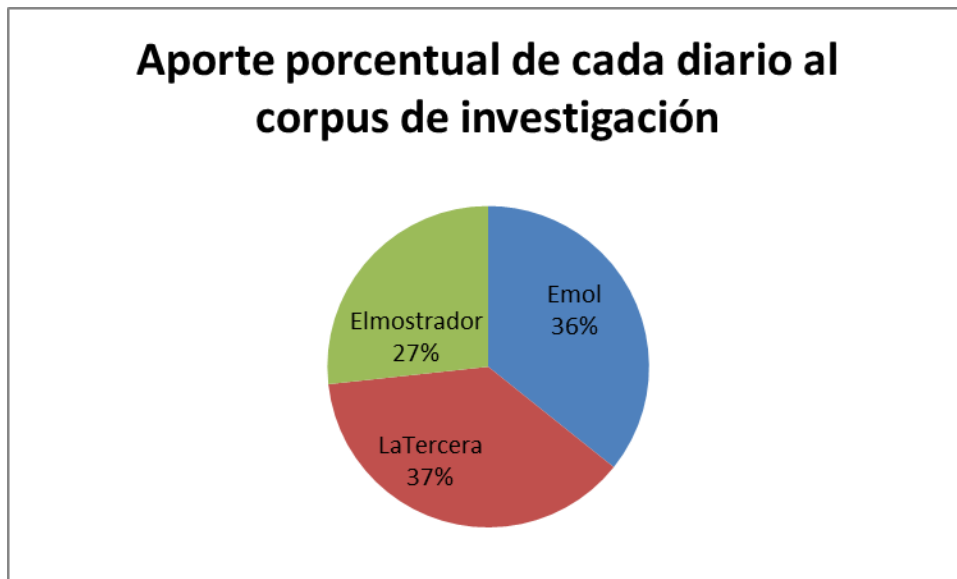


GRÁFICO 2 – Fuente: elaboración propia

El gráfico 3 muestra la contribución porcentual de cada diario online al corpus. La distribución no es homogénea ya que [www.elmostrador.cl](http://www.elmostrador.cl) solo aporta un 26%, lejos del 37% que aporta [www.latercera.cl](http://www.latercera.cl).

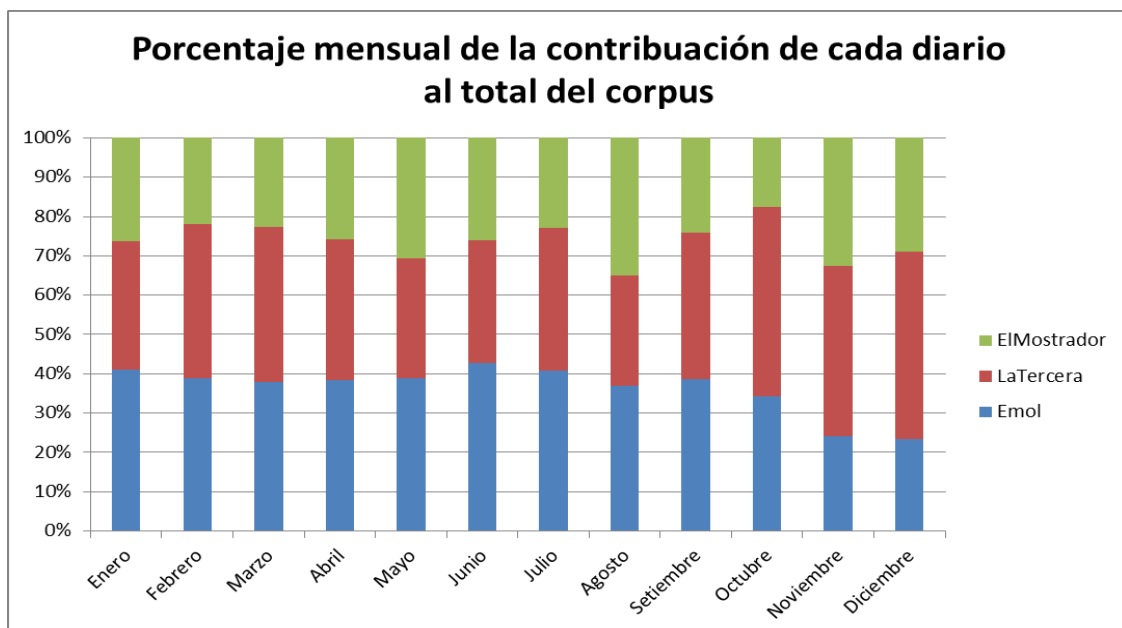


GRÁFICO 3 – Fuente: elaboración propia

Tal y como se puede observar en el gráfico 4, esta proporción porcentual tiende a mantenerse a lo largo de todo el año 2016, si bien en los meses de mayo y agosto, [www.latercera.com](http://www.latercera.com) queda relegada al tercer puesto en cuanto al porcentaje de noticias con las que contribuye al corpus, mientras que en los meses de noviembre y diciembre, es el [www.emol.com](http://www.emol.com) el que menos contribuye.

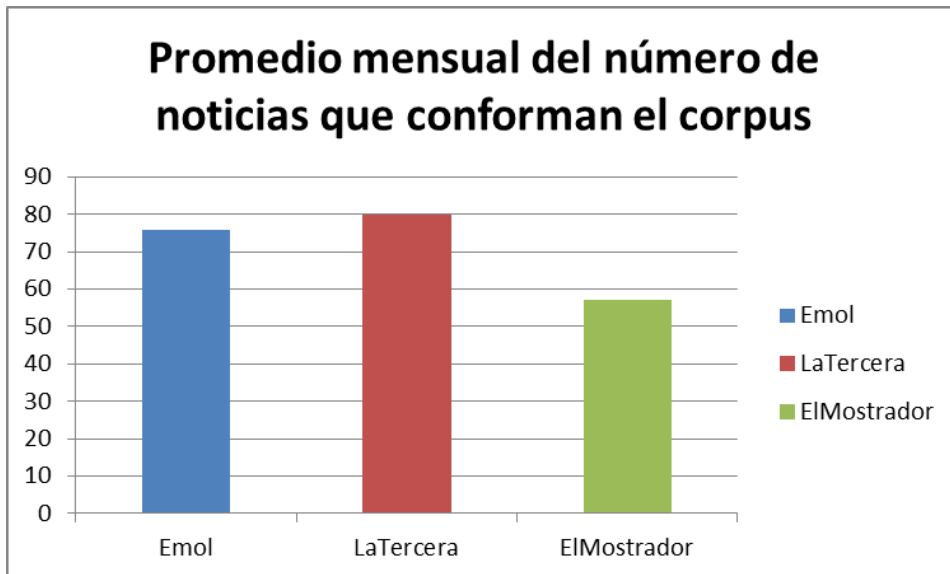


GRÁFICO 4 – Fuente: elaboración propia

El gráfico 5 muestra el promedio mensual del total de las noticias que los tres diarios habrían publicado durante 2016. Y el gráfico 6 muestra ese mismo promedio pero por día. Cuantitativamente, las diferencias entre los dos diarios online más leídos en Chile son insignificantes. En cambio, el [www.elmostrador.cl](http://www.elmostrador.cl) presenta un promedio mensual y diario significativamente inferior (- 26% aprox.).

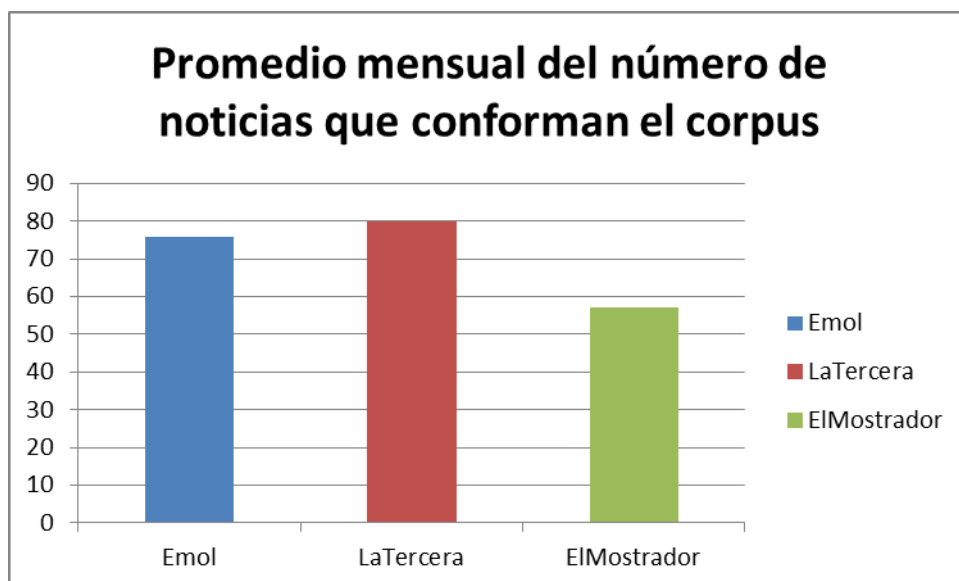


GRÁFICO 5 – Fuente: elaboración propia



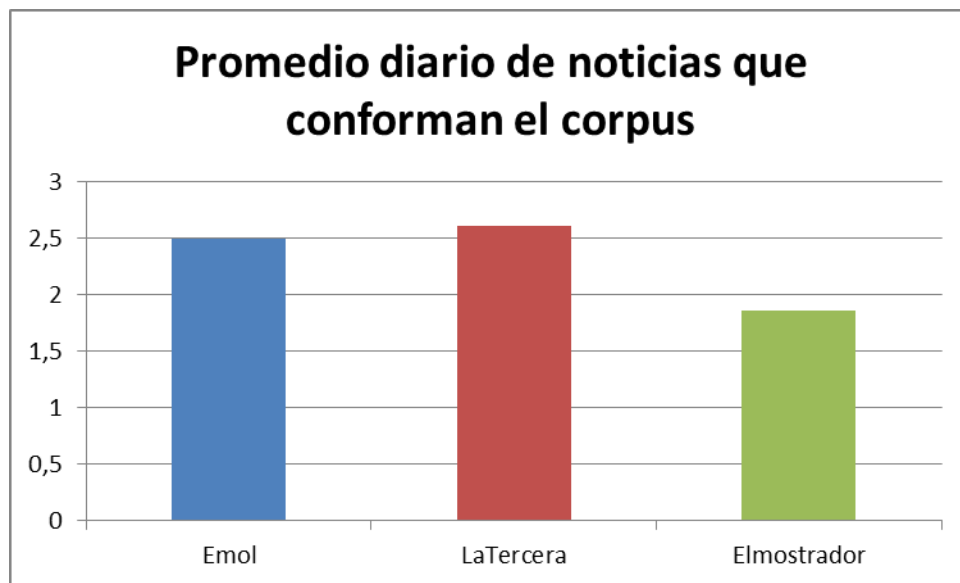


GRÁFICO 6 – Fuente: elaboración propia

#### 4. Discusión

A continuación, se comentan primero los aspectos de los resultados que pueden generar más dudas y, segundo, se discuten algunas de las posibilidades de explotación de los datos que hemos generado.

Llama la atención que el diario [www.elmostrador.cl](http://www.elmostrador.cl) cuente el mayor número de URLs al principio del proceso para terminar siendo el medio online que menos noticias aporta al corpus. Se ignora por qué Google arrojó más resultados para este diario que para el resto. En cualquier caso, esta diferencia inicial desaparece cuando se eliminan las URLs y los titulares duplicados, quedando el número de URLs igualado con los otros dos diarios.

La explicación para el gran número de URLs duplicadas en el caso de los tres diarios es simple: una misma noticia puede contener más de una palabra clave buscada. Por lo tanto, Google arroja una misma noticia en distintas búsquedas debido a que esta contiene alguna de las palabras clave solicitadas. Se estima que si se hiciera una búsqueda separada por cada una de las palabras clave, es decir, que no se hicieran búsquedas usando el boleano OR, el número de duplicados sería mayor.

También podría sorprender que los falsos positivos sumen dos tercios de las URLs arrojadas por Google (sin contar duplicados). Sin embargo, hay que tener en cuenta que la búsqueda que se realizó en Google con las palabras clave podría equipararse a la pesca de arrastre. Es decir, se pide a Google que arroje todas las URLs que contengan alguna de las palabras clave. Así, usando la palabra clave “cocina”, Google ofrecerá URLs en las que efectivamente la noticia sea sobre cocina o relacionado con ella, pero también URLs en las que la palabra “cocina” se emplee de modo metafórico, o anecdótico, o simplemente haya enlaces a otros contenidos que contengan la palabra cocina. De ahí que la codificación manual sea tan importante para poder separar los verdaderos positivos de los falsos.

Otro aspecto que es necesario comentar es la pequeña diferencia que se observa entre el número de URLs codificadas como pertinentes para el corpus y el número de noticias que finalmente consigue extraer import.io. Lo que hemos observado es que algunas URLs, o bien son casos que los

codificadores deberían haber rechazado por ser noticias en formato no textual, como audio, video, o fotografía, o bien son en realidad enlaces a documentos en formato .doc o .pdf.

Una vez discutidos los resultados parciales de cada fase del proceso de obtención de noticias que conforman el corpus, se comentan ahora los resultados finales, es decir, el número de noticias que cada diario aporta a la formación del corpus. Para empezar, es notable la similitud cuantitativa entre [www.emol.com](http://www.emol.com) y [www.latercera.com](http://www.latercera.com), y la diferencia entre estos dos diarios y [www.elmostrador.cl](http://www.elmostrador.cl). No debería sorprender que este último presente un 25% menos de noticias que los otros dos si se analizan estos tres medios online desde la Economía Política. Tal y como ya se ha descrito, en Chile hay una fuerte concentración de medios (Díaz, 2017) que acostumbra a calificarse de duopolio u oligopolio (Gronemeyer *et al.* 2013). Los dos grupos comunicativos principales son Mercurio SAP y Copesa, editores de [www.emol.com](http://www.emol.com) y [www.latercera.com](http://www.latercera.com) respectivamente. Estos dos grupos mediáticos concentraban en 2008 casi el 95% de la venta diaria de periódicos en papel y recibían el 65% de la torta publicitaria de la prensa (Mellado 2012). La situación actual, por lo menos en lectoría, es muy parecida (ACHAP AG, 2017). Así, mientras [www.emol.com](http://www.emol.com) y [www.latercera.com](http://www.latercera.com) son los dos diarios principales del duopolio mediático chileno con versiones en papel y online, [www.elmostrador.cl](http://www.elmostrador.cl) es un diario independiente editado por la empresa La Plaza S.A. que cuenta únicamente con una versión online y cuya lectoría es inferior a los otros dos diarios. Por lo tanto, desde un punto de vista cuantitativo, sería lógico que dos diarios generalistas con estructuras y poder económico similares produzcan, en términos absolutos, una cantidad de noticias sobre el hecho alimentario similar y superior a la que produce un diario generalista independiente con menos recursos económicos.

Si bien la cantidad de noticias presenta diferencias, se observa que la tendencia cuantitativa a lo largo del año de los tres diarios es bastante homogénea. En ocho de los 12 meses del año 2016 los tres diarios presentan una misma tendencia, y en los cuatro meses restantes, por lo menos dos de tres diarios siguen una tendencia similar. Quizá el caso más llamativo es el que se da en el mes de mayo, momento en que los tres diarios presentan un aumento significativo de noticias sobre el hecho alimentario. A falta de un análisis de contenido, un vistazo a los titulares de las noticias publicadas por los tres periódicos muestra que estos cubrieron ampliamente un evento crítico, a saber, la crisis de la marea roja que se dio en la isla de Chiloé, Chile, en mayo de 2016. Estas similitudes cuantitativas bien pueden ser un argumento más a favor de la fiabilidad del método aquí propuesto para conformar un gran corpus sobre el hecho alimentario. El hecho de ver tendencias muy parecidas, números de noticias semejantes, y que los tres hayan aumentado su publicación de noticias en mayo, demuestra que Google ha sido capaz de recuperar las noticias para todos los meses con una eficacia parecida. Por lo tanto, Google demostraría ser un buen árbitro a la hora de localizar las noticias sobre el hecho alimentario publicadas por distintos diarios. Por otro lado, y de nuevo a falta de un análisis de contenido, es muy probable que esta similitud sea también una prueba cuantitativa de la ya descrita homogeneidad de la agenda informativa en Chile (Díaz & Mellado, 2017).

En los gráficos 3 y 4 se puede observar en porcentajes el aporte total y mensual de cada diario al corpus. Llama la atención que si bien [www.elmostrador.cl](http://www.elmostrador.cl) es el que menos aporta —con una diferencia de 9,5 puntos porcentuales respecto a cada uno de los otros dos diarios— al total anual, en cuatro meses este mismo diario sea el segundo que más aporta. La explicación a esta diferencia habría que buscarla en la producción periodística de cada diario antes que en el método utilizado para localizar e identificar las noticias sobre el hecho alimentario. También podría ser un argumento más a favor de la consideración de largos periodos de tiempo en los estudios exploratorios interesados en caracterizar la cobertura del hecho alimentario.

Finalmente, los gráficos 5 y 6 muestran los promedios mensual y diario del número de noticias sobre el hecho alimentario identificadas para cada diario. A falta de estudios previos que hayan medido la cantidad de contenidos alimentarios publicados, solo queda la posibilidad de comparar dichos promedios con experiencia anecdótica de los autores. Una media de entre dos y tres noticias diarias no es menos de lo que uno podría esperar en prensa generalista, y sí, en cambio, probablemente algo más de lo que se esperaría a priori. Por consiguiente, el procedimiento aquí presentado permite asumir que la muestra obtenida es exhaustiva y altamente representativa, por lo menos en términos cuantitativos.

Más allá de los resultados, es necesario cerrar este apartado hablando de las distintas formas en que se pueden explotar las 2551 noticias identificadas y almacenadas en una hoja de cálculo. En primer lugar, existe la posibilidad de analizarlas usando distintos programas informáticos (software) diseñados expresamente para el análisis cuantitativo de texto. En *The Content Analysis Guidebook* (Neuendorf, 2016), hay una sección dedicada a la descripción de distintos software con los que extraer frecuencias, detectar coocurrencias, identificar relaciones estadísticamente significativas entre variables, términos, expresiones; además de analizar los contenidos a partir de diccionarios ya elaborados o construidos por el propio investigador. A partir de esta posibilidad de análisis, se abren por lo menos tres más.

La primera y más obvia tiene que ver con la posibilidad de realizar investigaciones con metodologías mixtas. Combinar análisis cuantitativo y cualitativo, incluso ayudado por software que permita trabajar cuantitativamente con categorías cualitativas construidas por el investigador, permite hacer investigaciones con grandes muestras y con un grado importante de profundidad.

La segunda posibilidad tomaría un primer análisis cuantitativo exploratorio como disparador de análisis mucho más reducidos en cuanto al número de casos, pero mucho más profundo en cuanto al estudio de los contenidos. Dicho de otra forma, en lugar de justificar investigaciones en comunicación y el hecho alimentario en teoría previa o en eventos críticos o emergentes, estos se podrían justificar en los resultados de un primer análisis cuantitativo. Por ejemplo, si un primer análisis cuantitativo arroja un porcentaje muy alto de noticias sobre alimentación y salud y un porcentaje muy bajo sobre el sector de la distribución alimentaria, sería necesario formular preguntas y objetivos de investigación que pudieran explicar dicha diferencia.

Finalmente, una tercera posibilidad sería usar el análisis cuantitativo para definir y delimitar muestras para estudiar una dimensión o tema concreto del hecho alimentario. En nuestro caso, estamos analizando la cobertura que hicieron estos tres diarios en 2016 sobre la relación entre alimentación y salud mental. Asistidos por un software se pueden identificar las noticias que trataron este tema y crear un nuevo corpus más reducido que analizaremos cuantitativamente y si lo estimamos oportuno, cualitativamente. Obviamente, este análisis también se podría complementar con un análisis crítico del discurso.

## 5. Conclusión

En este artículo se ha mostrado cómo es posible conformar de manera barata, efectiva, y técnicamente al alcance de cualquier persona con unos conocimientos informáticos básicos y unos mínimos conocimientos de inglés, corpus de miles de noticias publicadas en prensa digital que puedan ser analizadas con técnicas Big Data. La aplicación de este método se ha ejemplificado conformando un corpus de noticias sobre el hecho alimentario publicadas por tres diarios chilenos. Para ello, se ha partido de la teoría social para conceptualizar el hecho alimentario, y con la ayuda de la misma teoría junto con investigaciones recientes, se ha operacionalizado el hecho alimentario en un conjunto de palabras clave. Finalmente, con Google, [www.import.io](http://www.import.io) y tres codificadores, se ha obtenido un corpus de más de 2500 noticias sobre el hecho alimentario.

Los autores consideran que la comunidad académica de habla hispana debe incrementar la investigación sobre alimentación y comunicación porque los contenidos alimentarios que difunden los *mass media* contribuyen de manera significativa al conocimiento de la ciudadanía sobre el hecho alimentario. Además, estos mismos contenidos están reorganizando la relación de los seres humanos con la tierra y sus propios cuerpos. Analizar los contenidos alimentarios publicados por la prensa online es un primer paso primordial para analizar los efectos que estos pudieran tener en las relaciones que los humanos establecemos con el hecho alimentario. Y solo si se conceptualiza el hecho alimentario como un fenómeno social y humano total, y se operacionaliza como sistema alimentario y cultura alimentaria, conseguiremos abarcar la totalidad de los contenidos relacionados con la alimentación humana difundidos por los *mass media*.

Finalmente, los resultados muestran que, desde un punto de vista cuantitativo, la cobertura de los tres diarios chilenos estudiados es bastante homogénea, especialmente similar cuando se comparan los dos diarios con más audiencia online y que representan a los dos conglomerados de medios de comunicación que conforman el ya descrito por la literatura científica duopolio chileno. Estos dos diarios contribuyeron al corpus con un 10% más de noticias cada uno, en comparación con el tercer diario estudiado, que pertenece a un grupo de comunicación chileno independiente. La explicación para esta diferencia habría que buscarla en las diferencias de poder económico que habría entre este último diario y los dos pertenecientes al duopolio chileno, por un lado, y por el otro, a la homogeneidad de la agenda informativa en Chile.

- **Investigación financiada:** Este artículo es producto del proyecto de investigación titulado “Alimentación y prensa online: análisis cuantitativo de los contenidos alimentarios publicados por los cuatro diarios digitales chilenos más visitados.” (DI17-0089) financiado por la Universidad de La Frontera. La Universidad de la Frontera es una universidad pública de Chile que otorga fondos para la investigación mediante concurso interno. La evaluación de los proyectos de investigación que concursan es realizada por académicos externos a la Universidad de La Frontera.

Fechas:

- inicio de la investigación: junio 2017

- término de la investigación: noviembre 2018

## Notas

[1] Si bien no es incompleto, el esquema de sistema alimentario que presenta Goody es el más básico. Para una discusión sobre las diferentes definiciones sobre sistema alimentario que se han dado hasta el año 2005, véase Contreras, 2005. Para nuestros propósitos, consideramos que el esquema de Goody es más que suficiente.

[2] <http://www.alexa.com/topsites/countries/CL> [Consultado en julio de 2016 y en enero de 2019]

[3] Ver: <https://moz.com/blog/google-personalized-search> y <https://honeypotmarketing.com/remove-personalized-google-results/>

[4] A diciembre de 2018, este servicio permite 7 días de prueba con un máximo de 500 queries (consultas/búsquedas). Esto sería suficiente para hacer las pruebas oportunas y comprobar que el servicio es útil para el diario o diarios escogidos. La tarifa mínima es de 299 USD al mes con un máximo de 5000 queries. También está la opción de pedir un presupuesto para proyectos académicos que garanticen el número de queries necesario por un precio inferior al que se cobra a las empresas.

## Bibliografía

- ACHAP AG. (2017). Valida - Boletín de Circulación y Lectura 1º semestre 2017. Retrieved from <https://www.dropbox.com/s/xh7lk39edaaojro/Boletin%20de%20Circulaci%C3%B3n%20y%20Lectura%201%C2%B0%20semestre%202017.pdf?dl=0>
- Albala, K. (2013). *Routledge International Handbook of Food Studies*. Routledge.
- Alsina, M. R. (1989). *La construcción de la noticia*. Paidós España.
- Barthes, R. (1997). Toward a psychosociology of contemporary food consumption. *Food and Culture: A Reader*, 2, 28–35.
- Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.
- Bernabeu-Peiró, À. (2015). La divulgación radiofónica de la alimentación y la nutrición. El ejemplo de Radio 5 Todo Noticias. *Revista de Comunicación y Salud*, 5, 36–53.
- Blanco Hernández, N. (2015). La nota gastronómica y el artículo de costumbres. *Estudios Sobre El Mensaje Periodístico*, 21(2), 953–967.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard University Press.
- Bourdieu, P., Inda, A. G., Beneitez, M. J. B., Ordoñas, M. J. G., & Lalana, D. O. (2001). *Poder, derecho y clases sociales* (Vol. 2). Desclée de Brouwer Bilbao.
- Colle, Raymond. (2011). *El Análisis de Contenido de las Comunicaciones*. Sociedad Latina de Comunicación Social.
- Contreras, J., & Gracia Arnaiz, M. (2005). *Alimentación y cultura: perspectivas antropológicas*. Barcelona: Editorial Ariel.
- Cramer, J. M., Greene, C. P., & Walters, L. M. (Eds.). (2011). *Food as communication: Communication as food*. Peter Lang New York.
- Díaz, M. & Mellado, C. (2017). Agenda y uso de fuentes en los titulares y noticias centrales de los medios informativos chilenos. Un estudio de la prensa impresa, online, radio y televisión. *Cuadernos. Info*, (40), 107–121.
- Dijk, T. A. van. (2009). *Discurso y poder: contribuciones a los estudios críticos del discurso*.
- Douglas, M. (1972). Deciphering a meal. *Daedalus*, 61–81.
- Douglas, M. (1980). Las abominaciones del Levítico. *Pureza y Peligro. Un Estudio de Contaminación y Tabú*, 63–81.
- Duch, L., & Chillón, A. (2012). *Un ser de mediaciones*. Herder.
- Elias, N. (1989). *El proceso de la civilización. Investigaciones sociogenéticas y psicogenéticas*. fce.

Evans, J., Rich, E., Davies, B., & Allwood, R. (2008). *Education, disordered eating and obesity discourse: Fat fabrications*. Routledge.

Farré, M. (2004). *El noticiero como mundo posible: estrategias ficcionales en la información audiovisual*. La Crujía.

Frye, J., & Bruner, M. S. (2013). *The rhetoric of food discourse, materiality, and power*. New York: Routledge. Retrieved from [http://www.novanet.ebib.com/EBLWeb/patron/?target=patron&extendedid=P\\_1039373\\_0](http://www.novanet.ebib.com/EBLWeb/patron/?target=patron&extendedid=P_1039373_0)

Fúster, F., Ribes, M. Á., Bardón, R., & Marino, E. (2009). Análisis cuantitativo de las noticias de alimentación en la prensa madrileña en 2006. *Revista Española de Documentación Científica*, 32(1), 99–115.

Gaínza, G. (2003). La práctica alimentaria y la historia. *Lotman desde América*.

Goody, J., & Willson, P. (1995). *Cocina, cuisine y clase. Estudio de sociología comparada*. Editorial Gedisa.

Habermas, J., Domènech, A., & Grasa, R. (1981). *Historia y crítica de la opinión pública: la transformación estructural de la vida pública*. Gustavo Gili Barcelona.

Heldke, L. (2006). The Unexamined Meal is Not Worth Eating: Or, Why and How Philosophers (Might/Could/Do) Study Food. *Food, Culture and Society: An International Journal of Multidisciplinary Research*, 9(2), 201–219. <https://doi.org/10.2752/155280106778606035>

Koldobsky, Daniela. (2011). La gastronomía en el discurso crítico actual. *DeSignis*, (18).

Korthals, M., & Kooymans, F. (2004). *Before dinner: philosophy and ethics of food*. Dordrecht: Springer.

Krippendorff, K. (1990). *Método de análisis de contenido: Teoría y Práctica [Content Analysis Method: Theory and practice]*. Buenos Aires: Paidós.

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.

Lawrence, R. G. (2004). Framing obesity the evolution of news discourse on a public health issue. *The Harvard International Journal of Press/Politics*, 9(3), 56–75.

LeBesco, K., & Naccarato, P. (Eds.). (2008). *Edible ideologies: representing food and meaning*. Albany: State University of New York Press.

Lévi-Strauss, C. (1981). *El origen de las maneras de mesa* (Vol. 3). Siglo XXI.

Lévi-Strauss, C. (2013). The Culinary Triangle. *Food and Culture: A Reader*, 40.

Marín-Murillo, Flora, Armentia-Vizuet, José-Ignacio, & Olabarri-Fernández, Elena. (2016). Alimentación y Salud: Enfoques predominantes en la prensa española. *Revista Latina de Comunicación Social, En prensa*.

Martínez, A. A., & Poyatos, M. D. F. (2015). La gastronomía en la prensa española del siglo XIX/Gastronomy in the Spanish Press during the 19<sup>th</sup> century. *Estudios Sobre El Mensaje Periodístico*, 21(1), 17.

Menezes Ferreira, C., de Castro Oliveira, V. & Terrón Blanco, L. (2015). Una temática de peso: el tratamiento de la obesidad en los periódicos brasileños. In *La pantalla insomne*. Revista Latina de Comunicación Social. Retrieved from <http://www.revistalatinacs.org/15SLCS/libro-colectivo-2015.html>

Morin, E. (2014). *Le paradigme perdu. La nature humaine*. Seuil.

Narváez, T. A. (2013). Consejos dietéticos y nutricionales en la prensa española. *Revista Española de Comunicación En Salud*, 4(1), 17–26.

Nedelko, D. (2013, July 9). See How Your Website is Really Ranking - Unbiased Google Results. Retrieved January 2, 2019, from <https://honeypotmarketing.com/remove-personalized-google-results/>

Neuendorf, K. A. (2016). *The content analysis guidebook*. Sage.

Parasecoli, F. (2011). Savoring semiotics: food in intercultural communication. *Social Semiotics*, 21(5), 645–663. <https://doi.org/10.1080/10350330.2011.578803>

Plaza, J. F. (2012). Medios de comunicación, anorexia y bulimia. La difusión mediática del ‘anhelo de delgadez’: un análisis con perspectiva de género. *Revista Icono14. Revista Científica de Comunicación y Tecnologías Emergentes*, 8(3), 62–83.

Poulain, J.-P. (2017). *The sociology of food: eating and the place of food in society*. London; New York: Bloomsbury Academic.

Riffe, D., Lacy, S., & Fico, F. (2005). *Analyzing Media Messages-2/E.: Using Quantitative Content Analysis in Research*.

Roig, N. A. (2013). Alimentación y calentamiento global: «La larga sombra del ganado» en la prensa española. *Estudios Sobre El Mensaje Periodístico*, 19(1), 17–33.

Sánchez Gómez, F. (2010). La función didáctica del periodismo gastronómico en Internet. In *Alfabetización mediática y culturas digitales* (p. 172). Universidad de Sevilla.

Sandberg, H. (2007). *A matter of looks: the framing of obesity in four Swedish daily newspapers*.

Soler, J. (1997). The semiotics of food in the Bible. *Food and Culture: A Reader*, 55–66.

Strauss, C. L., Verón, E., & Menéndez, E. L. (1970). *Antropología estructural*. Eudeba.

Telfer, E. (2002). *Food for thought philosophy and food*. London; New York: Routledge. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=68639>

Thompson, J. B. (1998). *Los media y la modernidad: una teoría de los medios de comunicación*.

Thompson, J. R. (2012). Food talk: Bridging power in a globalizing world. *The Rhetoric of Food: Discourse, Materiality, and Power*, 58–70.

Velázquez, T. (1992). Los políticos y la televisión. Aportaciones de la teoría del discurso al diálogo televisivo. *Ariel, Barcelona*.

Weaver, D. A., & Bimber, B. (2008). Finding news stories: a comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, 85(3), 515–530.

Westall, D. (2011). La obesidad infantil en la prensa española. *Estudios Sobre El Mensaje Periodístico*, 17(1), 215–224.

Papers relacionados:

A C Yemsi-Paillissé, Y Acosta Meneses, M Martínez, E Calvo Gutiérrez (2018): “Aplicación de la crítica de los dispositivos a la cena performativa “El Somni” de El Cellar de Can Roca y Fran Aleu”. *Revista Latina de Comunicación Social*, 73, pp. 1267 a 1283.

<http://www.revistalatinacs.org/073paper/1306/65es.html>

DOI: [10.4185/RLCS-2018-1306](https://doi.org/10.4185/RLCS-2018-1306)

---

### Cómo citar este artículo en bibliografías / Referencia

R Sánchez Sabaté, C del Valle, M Mensa (2019): “Método para la construcción de grandes corpus temáticos de noticias de prensa digital. Hacia un corpus sobre el hecho alimentario”. *Revista Latina de Comunicación Social*, 74, pp. 594 a 617.

<http://www.revistalatinacs.org/074paper/1347/30es.html>

DOI: [10.4185/RLCS-2019-1347](https://doi.org/10.4185/RLCS-2019-1347)

### - En el interior de un texto:

R Sánchez Sabaté, C del Valle, M Mensa (2019: 594 a 617) ...

o

...R Sánchez Sabaté *et al*, 2019 (594 a 617) ...

Artículo recibido el 7 de diciembre. Aceptado el 23 de febrero.  
Publicado el 28 de febrero de 2019