



Revista de Comunicación

ISSN: 1684-0933

ISSN: 2227-1465

Universidad de Piura. Facultad de Comunicación

Mariela Lucina, Castro Cáceres; David Alejandro,
Chávez Salazar; Rubén, Urbizagástegui-Alvarado
Elecciones presidenciales en el Perú: minería de textos de los editoriales del diario La República
Revista de Comunicación, vol. 22, núm. 1, 2023, Marzo-Agosto, pp. 71-87
Universidad de Piura. Facultad de Comunicación

DOI: <https://doi.org/10.26441/RC22.1-2023-3025>

Disponible en: <https://www.redalyc.org/articulo.oa?id=589475109005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UPEM  redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Elecciones presidenciales en el Perú: minería de textos de los editoriales del diario *La República*

Presidential elections in Peru: text mining of *La República* newspaper publishers

Castro Cáceres, M. L., Chávez Salazar, D. A. y Urbizagástegui-Alvarado, R.¹

Recibido: 24-09-2022 – Aceptado: 2-02-2023

<https://doi.org/10.26441/RC22.1-2023-3025>

RESUMEN: Analiza los editoriales del periódico peruano *La República* publicados del 5 de mayo al 6 de junio de 2021, fechas en las que se llevaron a cabo las campañas políticas de la segunda vuelta electoral para ocupar la presidencia del Perú. Se pretende descubrir información relevante para entender el uso de las terminologías comunicadas a través del lenguaje escrito en ese periódico. Mediante la técnica de minería de texto y los algoritmos de aprendizaje automático para el análisis de los datos no estructurados, se creó un dendrograma y clústeres para validar y estimar las agrupaciones de los términos más frecuentes. En relación a cada candidato, el análisis de regresión muestra una asociación diferenciada en el uso de las palabras. Estas palabras conforman una red que expresa la coyuntura política del período en estudio y se refuerza con la construcción de una nube de palabras de mayor frecuencia de uso. Se concluye que los medios de comunicación construyen sus propias representaciones sociales sobre diversos temas locales y nacionales de modo que consciente o inconscientemente evidencian sus preferencias. Estas preferencias no son ajenas en los editoriales del diario *La República*. La importancia de realizar este tipo de estudio radica en el potencial que tiene la teoría de las representaciones sociales para colocar al investigador frente a la posición que tienen los sujetos, que construyen y comunican los embates antagónicos de la realidad social nacional.

Palabras clave: elecciones presidenciales; Perú; minería de textos; editoriales; diario *La República*.

ABSTRACT: It analyzes the editorials of the Peruvian newspaper *La República* published from May 5 to June 6, 2021, dates in which the political campaigns for the second round of elections for the presidency of Peru took place. The aim is to discover relevant information to understand the use of the terminologies communicated through the written language in that newspaper. Using the text mining technique and machine learning algorithms for the analysis of unstructured data, a dendrogram and clusters were created to validate and estimate the groupings of the most frequent terms. In relation to each candidate, the regression analysis shows a differentiated association in the use of words. These words form a network that expresses the political conjuncture of the period under study and is reinforced by the construction of a cloud of words with the highest frequency of use. It is concluded that the media construct their own social representations on various local and national issues in a way that consciously or unconsciously evidences their preferences. These preferences are not foreign in the editorials of the newspaper *La República*. The importance of conducting this type of study lies in the

¹ **Mariela Lucina Castro Cáceres.** Licenciada en Bibliotecología y Ciencias de la Información por la Universidad Nacional Mayor de San Marcos y Maestra en Educación con mención en Informática y Tecnología Educativa por la Universidad San Martín de Porres. Líneas de investigación: Análisis de la información, Organización de la información, Cienciometría. mlcastrocaceres@gmail.com, <https://orcid.org/0000-0003-0949-0352>

David Alejandro Chávez Salazar. Bachiller en Bibliotecología y Ciencias de la Información, Bibliotecólogo del Área de Apoyo a la Investigación en la Biblioteca de la Universidad de Lima. Líneas de investigación: Análisis de la información, Organización de la información, Cienciometría. dchavez@ulima.edu.pe, <https://orcid.org/0000-0002-7169-7672>

Rubén Urbizagástegui-Alvarado. Doctor en Ciencia de la Información por la Universidad Federal de Minas Gerais (Brasil), Maestro en Bibliotecología por la Universidad de Pittsburgh (EEUU), Bibliotecólogo en la Universidad de California en Riverside (EEUU). Sus principales líneas de investigación son: Análisis de la información, Sociología de la información y Cienciometría. ruben@ucr.edu, <https://orcid.org/0000-0001-5014-801X>

potential of the theory of social representations to place the researcher in front of the position held by the subjects, who construct and communicate the antagonistic onslaughts of the national social reality.

Keywords: Peru; presidential elections; text mining; editorials; La República newspaper.

1. Introducción

Las técnicas de minería de textos se definen como parte de la minería de datos que tiene como objetivo extraer conocimientos de datos textuales no estructurados (Gil, 2021). Los datos textuales *no estructurados* se refieren a la información almacenada en formato de textos que no siguen un patrón de organización predeterminado. Uno de los tipos más comunes de datos *no estructurados* es el que se genera y recopila en una variedad de formas, incluidos los documentos guardados en word, los artículos publicados en los diarios nacionales y/o locales, los mensajes de correo electrónico, las publicaciones de blogs, los sitios de redes sociales y otros. Los *datos estructurados* se refieren a la información que se encuentran en la mayoría de las bases de datos bibliográficas que ya tienen una estructura definida como autor, título, fuente, paginación, etc. Estos datos estructurados son archivos de textos que se suelen organizar y mostrar en filas y columnas, también pueden ser ordenados y procesados fácilmente por las herramientas de minería de datos.

En el panorama actual del llamado *big data*, en el que una gran cantidad de datos se generan diariamente, la aplicación de las técnicas de extracción automática de conocimientos se vuelven esenciales a partir de esos textos. Para la extracción automática de conocimiento en la minería de textos se utilizan algoritmos de aprendizaje automático (del inglés, *machine learning*) que se aplican a los textos no estructurados y que deben pasar por una serie de transformaciones orientados a la obtención de una representación estructurada. Normalmente, esta representación estructurada de los textos se configuran como vectores espaciales, formando una matriz de documentos por términos. En esta matriz, cada fila representa un documento y cada columna representa un término presente en la colección de documentos. Las palabras se utilizan comúnmente como atributos, dando lugar a una representación denominada *bolsa-de-palabras*. Este formato es simple y permite el uso directo del aprendizaje automático, porque asume que los términos son independientes y no se consideran las relaciones entre los términos (Sinoara *et al.*, 2021). Percibida de esta manera la minería de textos sería el proceso de analizar grandes cantidades de datos textuales a través de los avances logrados en el procesamiento del lenguaje natural que permitan el descubrimiento y extracción de algún aspecto pertinente dentro de una colección de textos no estructurados (Kumar, Kar & Ilavarasan, 2021). Una descripción de las etapas que lleva realizar la minería de textos puede ser consultada en Gaikwad *et al.* (2014) así como en Lanzarini *et al.* (2019).

La minería de textos no es un proceso de búsqueda de información, pues en ese tipo de búsqueda el usuario ya sabe lo que quiere, es decir, busca encontrar lo que quiere y/o necesita y lo encuentra. Ese es el proceso de recuperación de la información requerida y es el mecanismo utilizado en el análisis de textos estructurados. Ya en el análisis de textos no-estructurados, la minería de textos ayuda al usuario a descubrir informaciones desconocidas y a veces ni siquiera sospechadas por el propio usuario. Para ser más claros, la minería de textos, surgió de la necesidad de descubrir automáticamente información (patrones y anomalías) en los textos. El uso de esta tecnología permite recuperar información, extraer datos, resumir documentos, descubrir patrones, asociaciones y reglas, y realizar análisis cualitativos o cuantitativos sobre documentos de textos (Aranha y Passos, 2006; Gil, 2021; Urbizagástegui, 2021). El proceso de extracción de conocimiento implica entonces la evaluación e interpretación de los patrones o modelos obtenidos en la etapa de minería de datos, pues el lenguaje natural que las personas utilizan cotidianamente refleja de alguna manera las situaciones sociales y las relaciones interpersonales que se entablan con los demás agentes sociales; como la percepción y la visión de las contiendas cotidianas en la escena política y social de un país.

Por los motivos apuntados anteriormente, el presente artículo tiene como objetivo aplicar las técnicas de minería de textos para analizar los editoriales del periódico peruano *La República* publicados desde el 5 de mayo hasta el día 6 de junio de 2021, fechas en las que se llevaron a cabo las campañas políticas de la segunda vuelta electoral para ocupar la presidencia del Perú. Se pretende obtener información relevante orientada a entender el uso de las terminologías comunicadas a través del lenguaje escrito en ese periódico. La importancia de realizar este tipo de estudio radica en el potencial que tiene las técnicas de minería de textos para filtrar la visión que tienen los sujetos que construyen y comunican los embates antagónicos de la realidad social nacional. De ese modo, este estudio se centra en la búsqueda de la lógica del conocimiento práctico cuya forma de expresión empleada es el lenguaje objetivado, en este caso, en los editoriales publicados y comunicados por el periódico *La República*.

No se debe olvidar que “El editorial es el género periodístico que refleja más claramente la línea ideológica del periódico y así la hace pública” (Linares *et al.*, 2014, p. 144); y que además es:

Un espacio de debate mediatizado [que] se encuentra en las páginas editoriales de los periódicos nacionales [...]. En esta sección, la prensa escrita sostiene posiciones ideológicas e intenta influir en la agenda política. El editorial de un diario puede ser considerado como un género, ya que da cuenta de una manera particular de representar el mundo (discurso) y de configurar las identidades sociales (estilo). Específicamente, se trata de un texto sin firma a través del cual un medio de comunicación manifiesta su postura frente a temáticas de interés público, constituyendo un objeto de análisis de enorme valor para identificar las relaciones dialécticas que permiten el movimiento de significados y los procesos de mediación social”. (Cárdenas *et al.*, 2017, p. 21).

Citando diversas fuentes López (2016) sostiene que un editorial es un artículo que expresa la opinión de un periódico; explica que los editoriales presentan las opiniones de los periódicos como organizaciones colectivas y las opiniones de sus editores. Por lo tanto, puede generalizarse la idea de que los medios periodísticos suelen expresar una línea o tendencia editorial imparcial; sin embargo, como mencionan Núñez y Vázquez (2008)

La obligación de imparcialidad es relativa al tratamiento de la información, pero no es incompatible con la expresión de la opinión y, con ello, con la definición de una tendencia o línea editorial. La línea editorial o tendencia requiere una actitud uniforme en la toma de posición sobre los temas principales o de fondo. Hay que distinguir, pues, entre las distintas tendencias que el periódico puede adoptar o expresar a partir de los principios que se aplican al tratamiento de los temas de fondo, la elección política en el debate democrático y el comentario del acontecer. (p. 265).

Teniendo en cuenta que el impacto en la sociedad y la trascendencia práctica del oficio del periodismo en las sociedades democráticas está directamente ligado a las figuras de “libertad de opinión” y “el derecho a obtener una información veraz”, se exige que los conceptos se separen nítidamente, al menos, desde un punto de vista teórico” (Núñez, 2007). Sin embargo, el propósito de una editorial sería muy variado, ya que:

se define un punto de vista, se crea una opinión en el público sobre un suceso determinado, se analiza y se interpreta una noticia, se relaciona un suceso concreto con otros y se establecen juicios de valor que crean actitudes en una comunidad. (Moreno, 2003, p. 231).

Los medios de comunicación se han convertido pues, en un medio eficaz para influir en la ideología de las personas al dirigir la elección de las palabras y su presentación en el contenido. En el caso de la política, los medios de comunicación, especialmente los editores de los editoriales crean argumentos persuasivos que favorecen su postura política, plantean y defienden una tesis que transmite su postura sobre un tema controvertido y que suele estar relacionado con el interés público (Al Khatib *et al.*, 2016). La idea de que el periodismo no opera como un simple vehículo de información está siendo explorado con mayor énfasis en la literatura de las ciencias de la comunicación; esta idea

sugiere que el contenido ofrecido a las masas de lectores está influenciado por las preferencias políticas y el marketing, especialmente cuando se trata de los editoriales de los periódicos locales (Eberwein *et al.*, 2015).

Concordando con estas argumentaciones las preguntas de investigación que orientan este trabajo son las siguientes: ¿Cómo los editoriales de *La República* construyeron y difundieron los embates políticos en las últimas elecciones presidenciales?, ¿Fueron estos términos neutrales e imparciales?, ¿Puede la minería de textos ayudarnos a entender estos embates políticos?

Para responder a esas interrogantes y los objetivos propuestos, este documento está organizado de la manera siguiente: después de una breve introducción y establecimiento de los objetivos del artículo, se ofrece un marco teórico de la minería de textos aplicados a los estudios relacionados con los editoriales de un periódico o áreas cercanas a este asunto. Luego se describe la metodología, es decir, las unidades de análisis, la forma de recolección de los datos y la forma de medición de los mismos. Se presentan los resultados obtenidos y se exponen las conclusiones. Finalmente se lista la bibliografía que se empleó para la redacción de esta investigación.

2. Marco referencial

Las representaciones sociales son interpretaciones del mundo social que se elaboran colectivamente lo que es fundamental para que las personas pueden interactuar en el espacio social. Todos los aspectos que envuelven la vida de un sujeto, incluyendo el momento histórico-social en el que el sujeto está inmerso son formadores de las representaciones sociales que este mismo individuo se formulará respecto de los fenómenos sociales que forman parte de su contexto. Estas representaciones sociales se crean vía el proceso de anclaje y objetivación que sirven para darle familiaridad a lo desconocido; con lo cual, las representaciones sociales intervienen en una situación concreta en el que se sitúan los individuos y los grupos mediante la comunicación que se establecen entre ellos, por intermedio del marco conceptual proporcionado por su bagaje cultural, los códigos, valores e ideologías relacionadas con posiciones sociales ya establecidas. Las representaciones sociales son entonces producidos y formados en situaciones específicas (Moscovici, 1976, 2001, 2008). Este anclaje y objetivación se comunica a los demás vía el lenguaje oral o escrito, en el caso de los editoriales del diario *La República* de forma escrita, que es el objeto de análisis de este trabajo.

El análisis de textos utiliza técnicas del área del procesamiento del lenguaje natural, aplicadas a la minería de textos, que es una rama de la inteligencia artificial centrada en el diseño de métodos y algoritmos que toman como entrada y/o producen como salida datos en la forma de lenguaje humano. La exploración del procesamiento del lenguaje natural comienza aproximadamente por la década de 1950 con la llamada prueba de Turing que enuncia que si una máquina llegara a ser capaz de engañar a los seres humanos, haciéndose pasar por humana, con la misma facilidad con que un ser humano puede engañar a otro, entonces esa máquina tendría que ser considerarla como inteligente (Turing, 1950). Lo que se sugiere con esta afirmación es que debe existir alguna ligación entre las múltiples disciplinas para poder afrontar la tarea de comprender los procesos de comunicación humana y poder replicarlos en una computadora, pues el lenguaje natural se distingue de otros lenguajes artificiales por la riqueza de su vocabulario, así como por sus construcciones gramaticales y la multiplicidad de interpretaciones del sentido final del texto. Esta técnica de minería de textos está siendo explorada en muchos campos académicos y científicos.

Por ejemplo, Pollak *et al.* (2011), pensaron que una comparación de diferentes artículos periodísticos mostraría las discrepancias entre la cobertura noticiosa local (keniana) y la internacional (occidental), que puede explicarse en parte en términos de ideología. Para probar esta hipótesis, utilizaron la minería de textos para estudiar el uso del lenguaje en los informes periodísticos sobre

las elecciones de Kenia de diciembre de 2007 y la consiguiente crisis postelectoral. Tomaron sus datos de los editoriales de los periódicos kenianos *Daily Nation* y *The Standard* en contraposición a los periódicos británicos y estadounidenses *The Independent*, *The Times*, *The New York Times* y *The Washington Post*. La minería de textos encontró que la principal diferencia entre la prensa occidental y la local radicaba en el encuadre. En los periódicos británicos y estadounidenses, fueron creados un marco tribal mientras que los diarios de Kenia optaron por un marco sociopolítico. Los medios occidentales cuando explicaron los conflictos en Kenia vieron estos conflictos principalmente como una animosidad tribal o luchas por el poder entre tribus. Sin duda, muchos de los conflictos posteriores a las elecciones generales de 2007 tuvieron un aspecto étnico, pero siempre los conflictos sociales son más complejos.

Yuan (2016) intentó detectar y cuantificar los sesgos de los diarios en China. Para ese efecto, extrajo una matriz de términos versus documentos de los artículos sobre el XVIII Congreso del Partido Comunista Chino realizado en noviembre de 2012, de 21 periódicos chinos de siete provincias, así como de *el Diario del Pueblo*. Para esta matriz, utilizó la agrupación jerárquica para dividir los periódicos en dos grupos. Usando el dendrograma y las diferencias entre los grupos, pudo construir un índice para indicar la dirección y la magnitud del sesgo de los medios. En la muestra, los periódicos de *Zhejiang* y *Guangdong* constituyen un grupo y el resto constituyen otro grupo. La principal diferencia de los medios chinos se refleja en dos dimensiones: central/local y política/económica.

Nava y Marques (2019) realizaron un estudio longitudinal sobre la forma en que se construyó y representó la imagen pública de Lula Da Silva en los editoriales del diario *O Estado de São Paulo* durante las cinco campañas presidenciales en las que participó representando al Partido de los Trabajadores (años 1989, 1994, 1998, 2002 y 2006), para verificar si esta imagen cambió o se mantuvo igual a lo largo de las campañas. Los resultados del estudio evidencian que los términos más asociados y recurrentes a Lula resaltan características negativas, siendo relacionado a Brizola y a la extrema izquierda en la campaña de 1989 o asociado como un opositor menor a su competidor Fernando Henrique Cardoso en las elecciones de 1994, sin embargo, en 1998 ambos vuelven a enfrentarse siendo el foco de atención las propuestas económicas dada la coyuntura de crisis en el país. El estudio concluye que estos editoriales del diario *O Estado de São Paulo* se mostraban adversas a la candidatura de Lula Da Silva durante todas las campañas presidenciales, este posicionamiento casi no tuvo variaciones.

Para Bykov (2020), la “democracia y los derechos humanos” no configuran (temas/conceptos) clave en las políticas públicas de Rusia ya que los discursos presidenciales suelen centrarse en tres ejes comunes: Rusia, estado y Poder. En esta investigación el autor busca describir la evolución de la agenda política de una Rusia postsoviética y para ello analiza, a través de la minería de texto, los discursos políticos realizados por Boris Yeltsin (1994), Vladimir Putin (2000 y 2018) y Dmitry Medvedev (2008) a la Asamblea Federal de Rusia. Se sugiere que esta técnica puede ser aplicada para el estudio de textos políticos y automatizar investigaciones de lingüística política.

En una investigación sobre los editoriales de los periódicos en Japón realizada por Kaneko *et al.* (2021), se concluye que el análisis de texto asistido por computadora basado en el aprendizaje automático no supervisado, puede ser una técnica útil y en crecimiento para investigar diversos temas que usualmente aparecen en los medios de comunicación pero que se muestran ideológicamente divididos.

En el Perú, una descripción de las características discursivas de la línea editorial del diario *Correo* durante la segunda vuelta de las elecciones presidenciales del 2011, a partir del análisis de la sección central de 20 ediciones del diario, fue realizada por Suarez (2013). La observación del género de opinión en las páginas editoriales de los diarios *La República* y *Diario Uno* en la campaña política presidencial, del 1 al 15 de abril del 2016, fue desarrollada por Argumedo (2016). El tratamiento

informativo sobre la vacancia de Pedro Pablo Kuczynski en marzo del 2018 fue estudiado por Quispe (2019). También, un análisis morfológico y de contenido de 30 noticias de las ediciones digitales del diario *Expreso* y *La República* con relación al tratamiento periodístico sobre la disolución del Congreso en 2019 fue elaborado por Tineo (2020).

Como se puede apreciar en la literatura revisada, y hasta donde es de conocimiento de los autores del presente artículo, la minería de textos no ha sido tomada como herramienta para analizar los editoriales del diario *La República*.

3. Metodología

Este estudio se basa en la teoría de las representaciones sociales y se centra en la búsqueda de la lógica del conocimiento práctico, cuya forma de expresión es el lenguaje objetivado en los editoriales publicados y comunicados por el diario *La República*. Para la recolección de los datos, se copió cada editorial publicada en cada uno de los días del diario *La República* desde el 5 de mayo hasta el 6 de junio de 2021, día de las elecciones en segunda vuelta, y se guardaron en el folder creado con el nombre Republica. En total se copiaron y guardaron 33 editoriales, los cuales sirvieron como subsidios para el análisis de las palabras como representaciones sociales de conocimiento comunicados a los lectores del diario en mención. Luego se procedió a la preparación de los datos para que los algoritmos de la minería de textos no separen los nombres propios, por ejemplo, para que reconozcan nombres como Pedro Castillo y no Pedro por un lado y Castillo por el otro. Igualmente, Keiko Fujimori y no Keiko por un lado y Fujimori por otro lado. También para los nombres propios de todas las personas mencionadas en los editoriales, así como los nombres de los partidos políticos nacionales, las instituciones y las organizaciones.

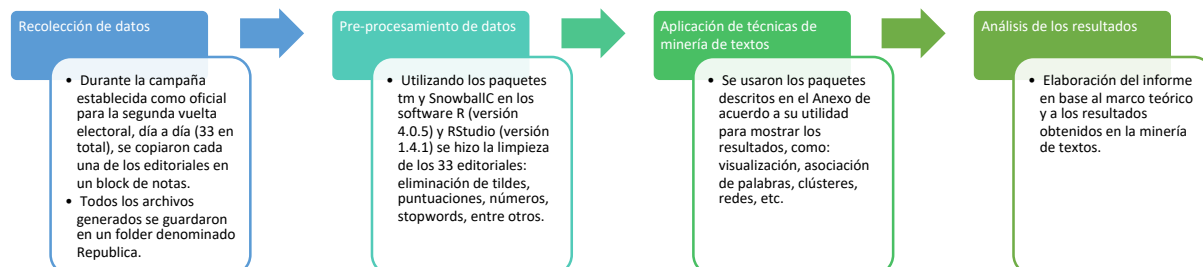
El software utilizado para análisis fue la versión 4.0.5 de R y 1.4.1 de RStudio, siendo elegidos porque son softwares que contienen recursos para el análisis de minería de textos y porque son gratuitos. Además, cuentan con un amplio abanico de funciones y se puede mejorar con el uso de nuevos paquetes, es decir, son programas potentes para el análisis de textos.

Los editoriales del periódico normalmente no se encuentran en un formato adecuado para la extracción de conocimiento, es necesario aplicar métodos de extracción e integración, transformación, limpieza, selección y reducción del volumen de estos datos, antes de la etapa de minería de textos. La integración significa obtener datos de los diversos editoriales (son 33 editoriales) y luego unificarlos, formando una única fuente de datos. La transformación es la idoneidad de los datos para su uso en algoritmos de extracción de datos. Luego viene la limpieza de la información que consistió en eliminar cualquier elemento que no sea texto (puntuaciones, fechas, números, etc.), eliminar espacios en blanco innecesarios y convertir todos los textos a minúsculas. Además, se eliminaron palabras que no aportan significado a los textos desde el punto de vista analítico (artículos, adverbios, etc.), así como algunas otras palabras que no aportan significados en este contexto en particular los llamados stopwords.

Todo este proceso fue realizado utilizando diversos paquetes de R (ver anexo) como *tm* y *readtext*. Para la remoción de los stopwords se utilizó el paquete *SnowballC*. Para la visualización de las frecuencias de los términos identificados se utilizaron los paquetes *ggplot2*. Estos mismos paquetes se utilizaron para visualizar las asociaciones o correlaciones entre los términos elegidos como interesantes. Ya para trazar las asociaciones entre las palabras se utilizaron los paquetes *graph* y *grid*. Para crear la nube de palabras se utilizaron los siguientes paquetes: *RColorBrewer*, *wordcloud* y *wordcloud2*. Los paquetes *fpc* y *cluster* se utilizaron para crear el dendrograma, validar y estimar las agrupaciones y finalmente realizar la agrupación de clústeres en torno a los términos más frecuentes. Este paquete contiene además funciones para el análisis de conglomerados y la medición de los datos; por eso se

usó en este trabajo para hacer un gráfico cusplot más ajustado a los datos. Finalmente, para trazar otros gráficos como barras, redes de las palabras más representativas utilizadas en los editoriales de *La República* se utilizaron los paquetes *igraph*, *textplot* y *glasso*. El proceso de organización, limpieza de los datos, así como el tratamiento de los textos están resumidos en la Figura 1.

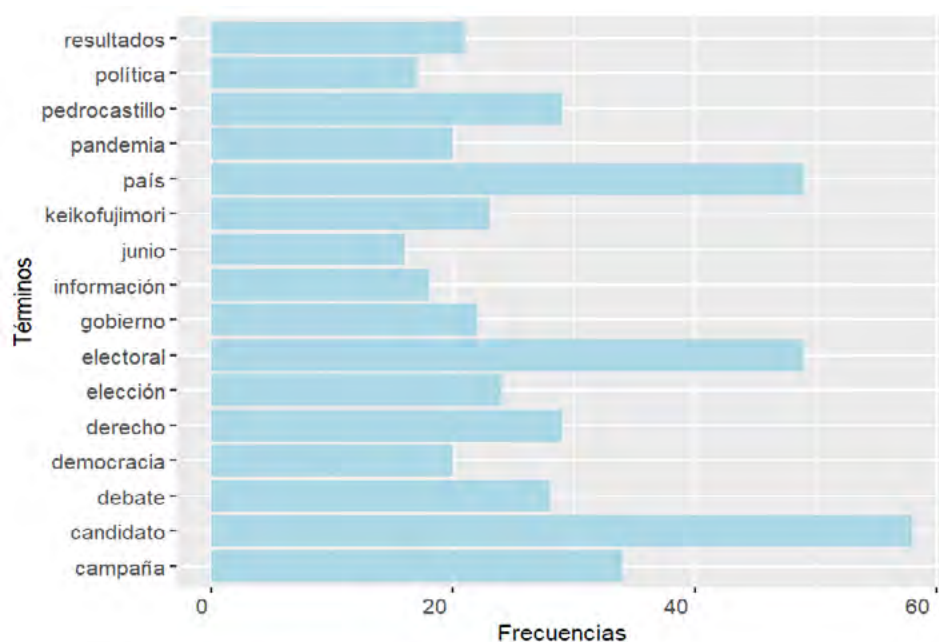
Figura 1. Tareas y herramientas empleadas en la investigación



4. Análisis de los resultados

En los 33 editoriales de *La República* se encontraron 3089 palabras diferentes. Para mostrar aquellas palabras cuya frecuencia de aparición fueron mayor a 10 veces, se construyó la Figura 2. Se aprecia que “candidato” (1.8%), “electoral” (1.6%) y “país” (1.6%) son las palabras más utilizadas en el periodo de estudio.

Figura 2. Frecuencia de palabras más utilizadas en los editoriales



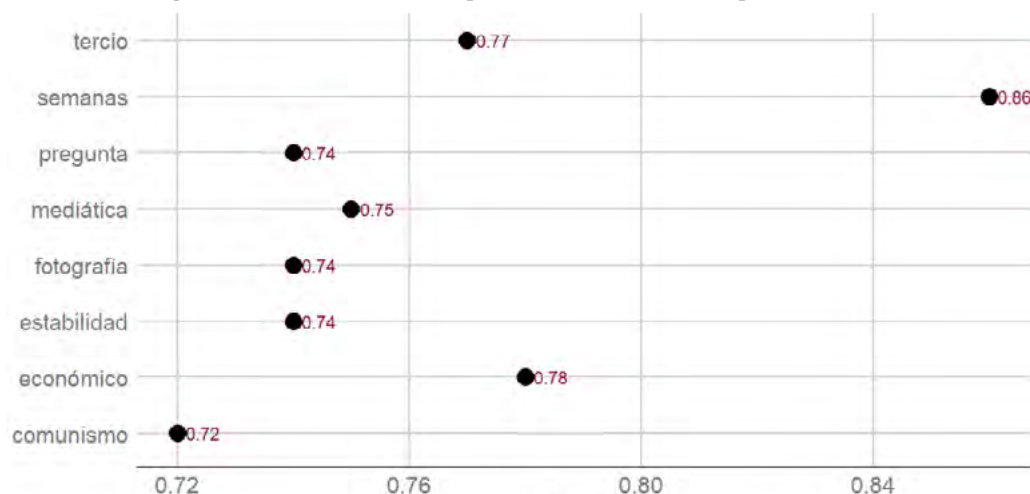
Igualmente las palabras “keikofujimori”, “pedrocastillo”, “resultados”, “elección”, “democracia”, etc., tienen coherencia con la coyuntura política del momento y muestran la representación mental que el editorialista de *La República* tenía sobre los embates políticos del periodo en estudio. Sin embargo, no es posible olvidar que en la práctica social, esas representaciones mentales están orientadas a la producción de efectos sociales; a producir y reproducir lo que aparentemente designan como una realidad objetiva pero son apenas representaciones mentales de actos de percepción y apreciación en la que el editorialista invierte sus intereses y estrategias de manipulación simbólica que funcionan como signos, emblemas o estigmas pero también como poderes políticos.

La extracción de asociaciones de palabras (ver la Figura 3 y 4) se basa en las relaciones gramaticales entre las palabras presentes en los textos que hace uso de las dependencias temáticas para expresar e identificar informaciones referenciales entre argumentos sintácticos. Sirve para medir las relaciones entre las palabras presentes en los 33 editoriales de *La República* develando patrones de percepción inherentes a las estructuras semánticas que emergen de la relación entre las principales unidades léxicas extraídas. De esa manera se conforman descriptores con alto grado de asociación entre sí, representando las inquietudes centrales de la percepción del editorialista acerca de la coyuntura política vivida en el momento.

Estas asociaciones de palabras pueden mostrar una asociación positiva, negativa o neutral. Para mostrar con qué palabras se asocia al candidato Pedro Castillo se construyó la Figura 3. A un nivel de correlación de Pearson positiva mayor a 0.7, se le asocia al candidato con la palabra “semanas” ($r = 0.86$), refiriéndose a las pocas, escasas semanas que quedan para el día de las elecciones y a las semanas previas para ese día. La palabra “tercio” ($r = 0.77$) se refiere a que casi un tercio de los votantes oscila entre no acudir a votar, votar en blanco o votar nulo, pero que al mismo tiempo si acuden a votar un tercio de la población lo haría porque no quieren que el fujimorismo vuelva al poder; es el llamado antivoto que persigue al fujimorismo.

Las palabras “mediática” ($r = 0.75$), “pregunta” ($r = 0.74$) y “fotografía” ($r = 0.74$) son asociados a Pedro Castillo debido a que las encuestas de opinión solían preguntar por los motivos que un desconocido, que utilizaba un lápiz como símbolo de campaña y que al inicio no aparecía en los primeros lugares, logró llegar a la segunda vuelta frente a una rival con una fuerte operación mediática, con bombardeo masivo en paneles y redes sociales; esta derrota paulatina conforma una fotografía del momento. Sin embargo, lo más destacable de estas asociaciones giran en torno a las palabras “económico” ($r = 0.78$), lo cual era una de las preocupaciones centrales respecto a su candidatura, ya que el candidato se relacionaba con el “comunismo” ($r = 0.72$), lo cual traería como consecuencia una falta de “estabilidad” ($r = 0.74$) en diversos planos.

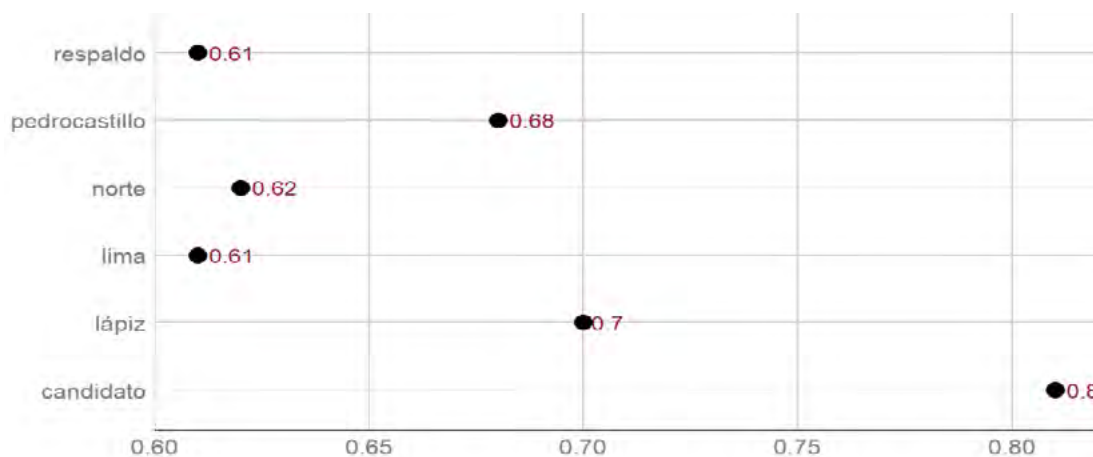
Figura 3. Asociaciones de palabras con término “pedrocastillo”



Para el caso de la candidata Keiko Fujimori (Figura 4), las asociaciones se determinaron a nivel de correlación de Pearson positiva mayor a 0.6, resultando que las palabras con mayor nivel de asociación son “candidato” ($r = 0.8$), seguida de “lápiz” ($r = 0.7$) y muy cerca también “pedrocastillo” ($r = 0.68$); es decir, *La República* muestra a Keiko Fujimori como “candidato” y la relaciona con su contendor Pedro Castillo cuyo símbolo era el “lápiz”. La palabra “respaldo” se refiere a las organizaciones políticas que anunciaron el apoyo a la candidatura del fujimorismo; a las encuestas de opinión donde los sectores socioeconómicos D y E (medio bajo y bajo) respaldan mayoritariamente

la candidatura del fujimorismo; se sostiene que los candidatos tendrán que convencer a aquellos que ya decidieron su voto, pero que aún pueden cambiar de opinión para que los respalden. Las palabras “norte” ($r = 0.62$) y “Lima” ($r = 0.61$) se refiere a las macrozonas donde la candidata Keiko Fujimori obtiene mayor respaldo: en el norte del país y en el norte de Lima.

Figura 4. Asociaciones de palabras con término “keikofujimori”



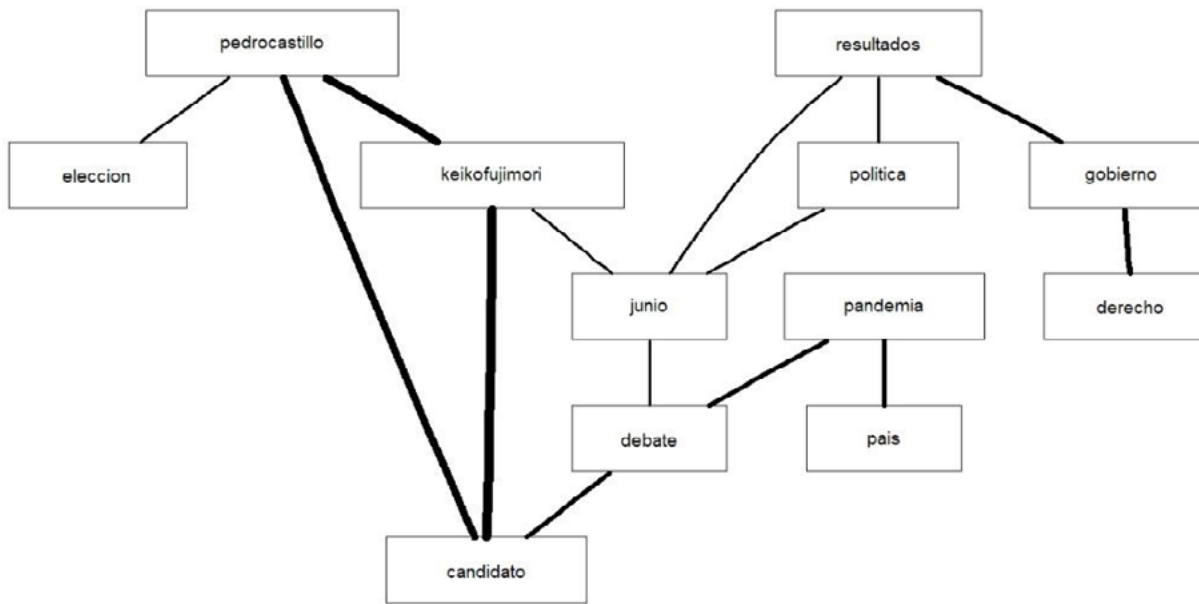
Se puede observar una diferencia entre las palabras que se asocian a uno y otro candidato. Mientras que para Pedro Castillo se presentan temas específicos del quehacer político y social, a la candidata Keiko Fujimori se le asocia con su contendor y el respaldo que recibe en determinadas regiones del país, mostrando así un tratamiento diferenciado y venal.

Es preciso señalar que en el caso de Pedro Castillo se tomó (convenientemente para el estudio) una correlación de Pearson 0.7 ya que en este nivel la correlación mostraba asociaciones con una cantidad de términos adecuados para el análisis. Si se tomaba una correlación mayor a 0.6 las asociaciones aumentaban a un elevado número de términos poco útiles para el análisis. Para el caso de Keiko Fujimori, cuando se usó un nivel de 0.7 la cantidad de términos asociados mostraba apenas un término, por eso fue más conveniente para el análisis usar una correlación mayor a 0.6. Es por esta razón, que no visualizamos en la Figura 3 la correlación entre ambos candidatos a pesar que esta se evidencia en la Figura 4.

Se denomina red semántica al esquema que permite representar la estructura de la red de palabras mediante un gráfico donde se representan los nodos y lazos de las palabras mostrando además la significación de los conceptos relacionados. Los conceptos, expresados a través de la palabra, surgen en un contexto social histórico en la forma de representaciones mentales de significados siendo parte del sujeto que la emite como un ser que está en la historia y en un espacio social cambiante. Ese espacio social es el lugar donde las relaciones sociales se articulan con saberes, visiones, percepciones, sueños o ficciones del emitente. Es en este espacio complejo donde se realiza el poder que el lenguaje representado ejerce sobre los agentes en comunicación y se busca la transformación en algo que otorga nuevas posibilidades de saberes, visiones, percepciones de tal manera que este modela una especie de compromiso con la acción transformadora de la sociedad.

La Figura 5 muestra la red de palabras de los editoriales de *La República* y sus vinculaciones, siendo el valor 0,24 para determinar las redes de relaciones de las palabras más utilizadas. El grosor de las líneas que unen a las palabras indica una mayor o menor relación. El centro de las preocupaciones de los editoriales del diario *La República* giran en torno a Keiko Fujimori y Pedro Castillo como candidatos, lo que expresa la coyuntura política del momento.

Figura 5. Redes de relaciones de las palabras más utilizadas en los editoriales



Por otro lado, se observa una relación fuerte entre ambos candidatos, así como entre las palabras “resultados” y “gobierno”. En menor medida, se aprecia una relación entre las palabras “candidato”, “debate”, “pandemia” y “país”, lo cual tiene bastante significancia ya que es uno de los temas presentes en el discurso. También podemos apreciar una relación fuerte entre “candidato”, “debate” y “pandemia”, “país” y “pandemia”, lo cual tiene bastante significancia ya que era uno de los temas presentes en los debates entre los candidatos. Obsérvese que la palabra “elección” sólo está relacionada con “pedrocastillo”. La palabra “junio” funciona como puente en esta red, es una palabra clave en la red, porque opera como mecanismo de comunicación con todas las otras palabras de la red. Junio es también el mes de las elecciones, por eso su importancia en la red.

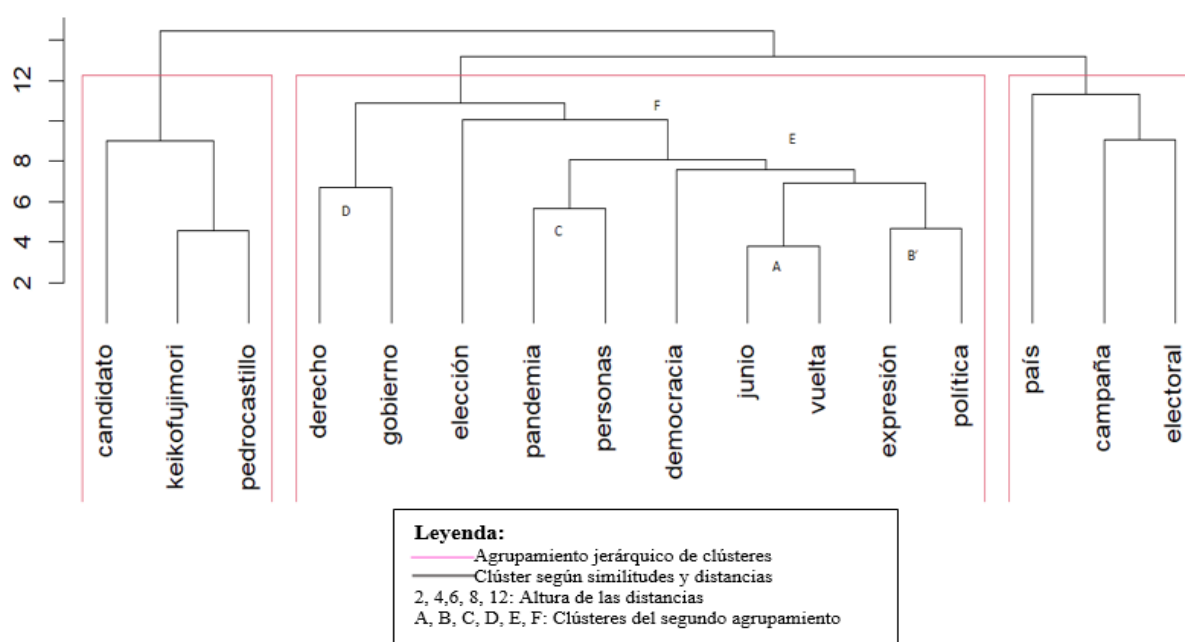
Figura 6. Nube de palabras y nube de bigramas



Se configuró el programa para que muestre una nube de palabras, con un máximo de 30 palabras que estén presentes en el corpus con una frecuencia mínima de 3 veces, la variación en el tamaño y el color de las palabras indican su mayor frecuencia. Se observa que las palabras “candidato” y “elección” son las que más se repiten en los editoriales, mientras que “educación”, “peruanos”, “expresión”, “político”, “voto”, “vuelta” y “presidencial” aparecen con el mismo color y en menor tamaño, es decir, son las que se presentan con menor frecuencia. R también nos permite hacer bigramas y trigramas de acuerdo a las necesidades del investigador y mostrarlas en barras o nube de palabras como se muestra en la Figura 6. Para realizar la nube de palabras de bigramas, es decir, grupos de dos palabras, se configuró el corpus con una frecuencia de 15 y que muestre un máximo de 36 grupos de palabras. Resaltan, “proceso electoral” y “segunda vuelta”, seguido de “campana electoral” y “candidato pedrocastillo”. Ello tiene una gran concordancia con la coyuntura del periodo en estudio.

El dendrograma es una representación gráfica de estructura arborescente de dependencia que muestra las distancias y las similitudes entre elementos. Agrupa en base a la similitud de sus características a dos individuos en un grupo, también llamado clúster. Los diversos clúster se van formando hasta formar uno solo que contiene la totalidad de los mismos. Así mismo, muestra la distancia entre los clúster. Los clústeres se representan mediante líneas horizontales, mientras que las líneas verticales (altura) representan las distancias, es decir, la altura en que se unen dos grupos es la distancia entre ellos, por tanto, a mayor altura, mayor diferencia. También, si el investigador desea, se pueden formar determinados número de grupos, los cuales se muestran en forma rectangular. La Figura 7 muestra el dendrograma de tres agrupamientos jerárquicos de las palabras más utilizadas en los 33 editoriales de *La República*.

Figura 7. Dendrograma de las palabras usadas con mayor frecuencia



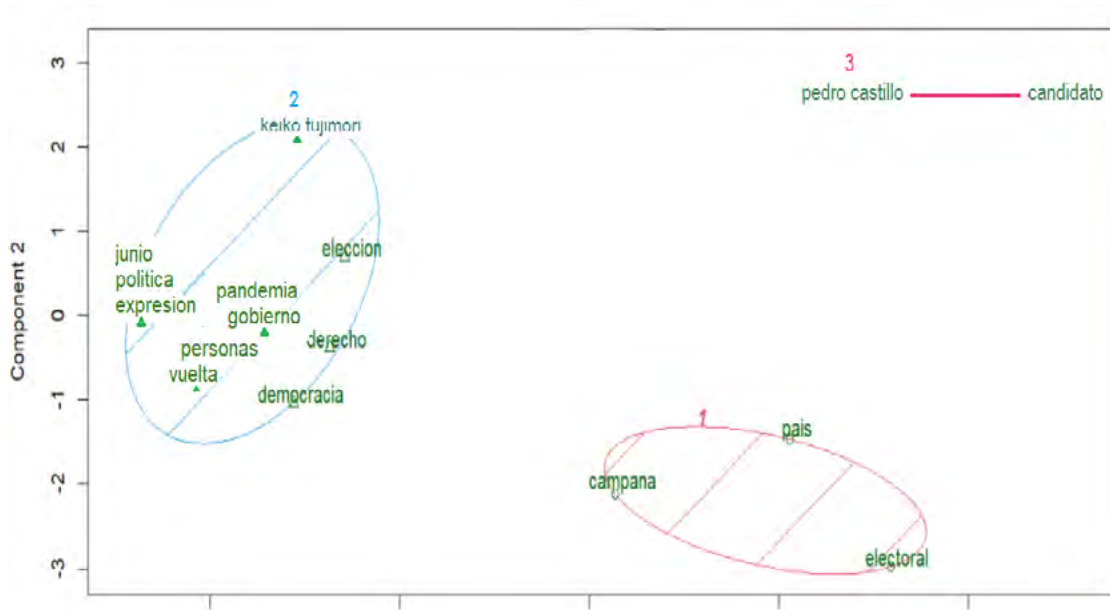
En el primer agrupamiento lo componen el clúster “keikofujimori” y “pedrocastillo” con “candidato”. El segundo agrupamiento, se observa en un primer nivel el clúster (A) formado por “junio” y “vuelta”. En un nivel inmediatamente superior, se unen “expresión” y política (clúster B) seguidos por “pandemia” y “personas” (clúster C) y finalmente, “derecho” y “gobierno” (clúster D). Un paso siguiente engloba el clúster A y B con “democracia” (clúster E). Siguiendo las agrupaciones, el clúster E junto con el C y la palabra “elección” forman el clúster F. Finalmente, el clúster D

y F forman este segundo gran agrupamiento. El tercer agrupamiento está conformado por el clúster formado por las palabras “campana” y “electoral” más la palabra “país”.

Al igual que en los otros gráficos, se aprecia una coherencia entre los términos agrupados con la coyuntura del periodo en estudio. En los grupos extremos presentan, por un lado a “keikofujimori” y “pedrocastillo” con la palabra “candidatos”, efectivamente los dos estaban en disputa; en el otro extremo, las palabras “campana”, “electoral” y “país”, que representa como uno de los temas relevantes de aquel momento. El agrupamiento central, también se aprecia una coherencia como en el clúster A con las palabras “vuelta” y “junio”, ya que en este mes se realizó la segunda vuelta que definió al presidente. Igual, el caso de “pandemia” y “personas”, pues eran los temas de preocupación; así sucesivamente podemos observar las relaciones.

En la figura 8, se muestran los tres componentes identificados por el método de Ward. Estos representan el 95% de la variabilidad de los datos. Se aprecia que “pedrocastillo” y “candidato” son las palabras (variables) más representadas y con un alto rango por el tercer componente. Del mismo modo, “keikofujimori” es la palabra más representativa y con un alto rango por el segundo componente. Sin embargo, las otras palabras que conforman el segundo componente (“elección”, “derecho”, “gobierno”, “pandemia”, “personas”, “democracia”, etc.) así como, las palabras “país” “campana” y “electoral” del primer componente, tienen baja proporción y representatividad en los editoriales del diario *La República*.

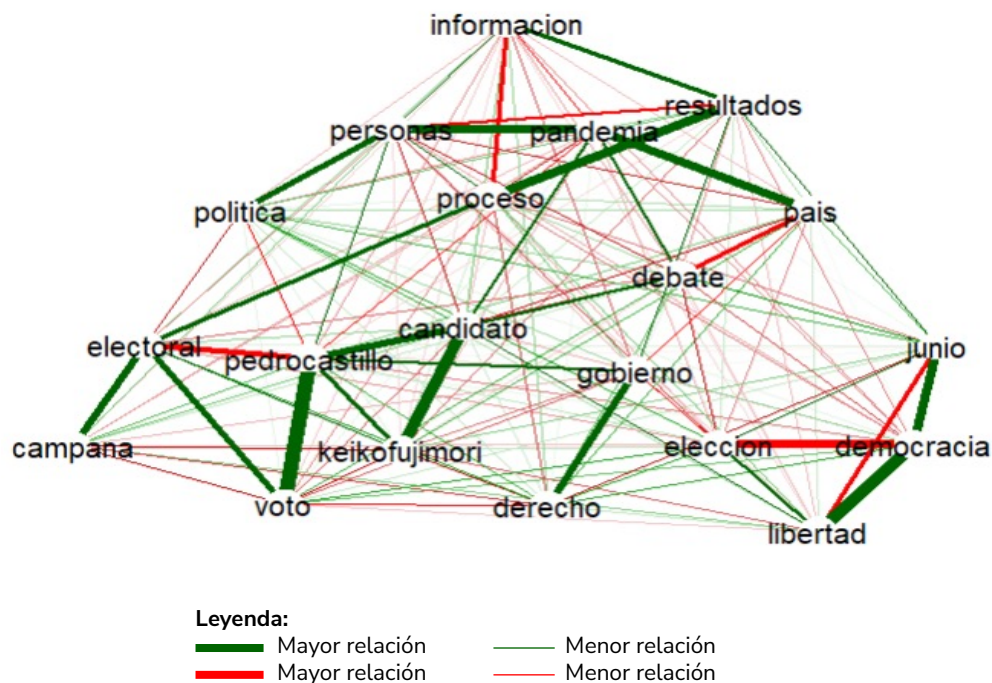
Figura 8. Clústeres de las palabras más utilizadas



Uno de los temas centrales del análisis de textos es la identificación de los conceptos de interés y la identificación de las relaciones que se producen entre estos conceptos lingüísticos. En la Figura 9 se aprecia las diferentes relaciones entre las 20 palabras más representativas rescatadas de los editoriales. Se observa que la palabra “candidato” tiene una relación mayor con “keikofujimori” -se define por el grosor de la línea-, también se relaciona con “pedrocastillo”, lo cual se asemeja a lo obtenido en las asociaciones de palabras con el término “keikofujimori” (Figura 4), donde “candidato” tiene un mayor nivel de asociación, a su vez está con “debate”, conformando un espectro de preocupaciones de la escena política nacional vivida en el momento del proceso electoral. Otra relación fuerte se aprecia entre las palabras “gobierno” y “derecho” así como entre “proceso” y “resultados” y entre “democracia” con “libertad”. Otro grupo se centra en los resultados del proceso, que está

vinculado a las elecciones del mes de junio. Este gráfico nos demuestra que en los editoriales de *La República*, durante el periodo de estudio, el tema principal fue la segunda vuelta electoral para elegir nuevo presidente.

Figura 9. Redes de palabras de los términos utilizados en los editoriales



5. Discusión y conclusiones

En breve tiempo y con bajo costo -ya que se utilizó software libre Rstudio-, se analizaron los editoriales del diario *La República* publicados del 5 de mayo al 6 de junio de 2021, es decir, durante la campaña electoral hacia la segunda vuelta electoral para la elección presidencial. Los 33 editoriales en estudio produjeron 3089 términos con una dispersión del 95% y la palabra de mayor longitud tenía 28 letras. Mediante las diversas técnicas de minería de textos se ha podido obtener los términos utilizados y sus relaciones, lo cual es importante ya que “quien ignore los principios editoriales de la empresa informativa y no sepa qué persigue o pretende puede ser juguete en manos del informador; y el desconcierto o la ignorancia serán el resultado de unos principios editoriales no formulados con claridad” (Edo, 1994, p. 170). En los primeros resultados del análisis se observa que las palabras que aparecen con mayor frecuencia y que a su vez tienen mayor relación son “pedrocastillo”, “candidato” y “keikofujimori”. A Pedro Castillo se le presenta con temas específicos del quehacer político y social mientras que a la candidata Keiko Fujimori se le asocia con su contendor y se le confiere un tratamiento distinto. Los términos más utilizados están relacionados a la coyuntura electoral, como son “candidato” y “elección”, términos que más se repiten en los editoriales, así como “proceso electoral” y “segunda vuelta”, que de hecho coinciden plenamente con el momento político efervescente que se vivía en esos días.

Pedro Castillo aparece como candidato del partido político Perú Libre cuyo símbolo es el lápiz pero no sucede lo mismo con el partido Fuerza Popular y su candidata Keiko Fujimori. Esta mención puede deberse a la importancia que le dan los medios de comunicación al partido Perú Libre y a lo que representa como ideología de extrema izquierda. Esto se ve reflejado también en los resultados obtenidos de las asociaciones de la palabras donde se aprecia una relación entre “pedrocastillo” y las palabras “comunismo” a un nivel del 0.72%, “estabilidad” (0.74%) y “económico” (0.78%).

Considerando que, “los medios de comunicación masiva elaboran sus propias representaciones sociales sobre diversos temas, hechos o personas y es el sujeto o grupo al que se pertenece, expuestos a los medios de comunicación, los que pueden rechazarlas o incorporarlas” (Cuevas, 2011, p. 5) y teniendo en cuenta que a los medios de comunicación se les considera como el *cuarto poder* por su influencia en los asuntos sociales y políticos, se observa cómo en los editoriales del diario *La República* se presenta a “keikofujimori” como “candidata” con “respaldo”, mientras que a su contendor “pedrocastillo” se le relaciona más intensamente con las palabras “económico”, “comunismo” y “estabilidad” pero sugiriendo más bien inestabilidad económica. Es decir, se representan de diversa manera a cada uno de los candidatos, lo que pudo haber influenciado en la reacción de los votantes.

Las aparentes verdades políticas, especialmente en un momento crítico como son las elecciones políticas, lleva a olvidar que los productos lingüísticos son también representaciones mentales específicas de este espacio social particular que es el periodismo, y que esas representaciones lingüísticas que son los editoriales cumplen funciones políticas, conscientes o inconscientes, y evidencian sus preferencias. Este es el caso del diario *La República*.

Bibliografía

- Al Khatib, K., Wachsmuth, H., Kiesel, J., Hagen, M., y Stein, B. (2016, December). A news editorial corpus for mining argumentation strategies. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 3433-3443. <https://aclanthology.org/C16-1324/>
- Aranha, C. y Passos, E. (2006). A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, 5(2), 1-8. <https://doi.org/10.21529/RESI.2006.0502001>
- Argumedo Solano, L. (2016). *Análisis comparativo de la presentación del género de opinión en la página editorial de los diarios “La República” y “Diario Uno” en la campaña política a la presidencia de la República, 1 al 15 de abril, 2016* [Tesis de Licenciatura, Universidad César Vallejo, Lima, Perú]. <https://hdl.handle.net/20.500.12692/2350>
- Benoit, K. y Obeng, A. (2021). *Readtext Import and Handling for Plain and Formatted Text Files*. R package (version 0.81) [Software]. <https://CRAN.R-project.org/package=readtext>
- Bouchet-Valat, M. (2020). *SnowballC: Snowball Stemmers Based on the C ‘libstemmer’ UTF-8 Library* [Software]. <https://cran.r-project.org/web/packages/SnowballC/index.html>
- Bykov, I. A. (2020). Studying political discourse of the President address in Russia with the text mining technique. *Journal of Philosophy, Culture and Political Science*, 3, 68-75. <https://doi.org/10.26577/jpcp.2020.v73.i3.08>
- Cárdenas Neira, C., Cabalin, C. y Montero, L. (2017). Análisis crítico de editoriales y su influencia en la conformación de políticas sobre la calidad de la educación en Chile: Los casos de La Tercera y El Mercurio. *Estudios pedagógicos (Valdivia)*, 43(1), 19-39. <https://dx.doi.org/10.4067/S0718-07052017000100002>
- Csardi, G. y Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1-9. https://www.researchgate.net/publication/221995787_The_Igraph_Software_Package_for_Complex_Network_Research
- Cuevas, Y. (enero-junio, 2011). Representaciones sociales en la prensa: aportaciones teóricas y metodológicas. *Sinética*, 36, 1-19. <http://www.scielo.org.mx/pdf/sine/n36/n36a3.pdf>
- Eberwein, T., Porlezza, C. y Splendor, S. (2015). Media as political actor. En G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication*, 703-711. Wiley-Blackwell.
- Edo, C. (1994). *La crisis de la prensa diaria: la línea editorial y la trayectoria de los periódicos de Madrid*. Ariel.
- Feinerer, I., Hornik, K. (2020). *tm: Text Mining Package*. R package (version 0.7-8) [Software]. <https://CRAN.R-project.org/package=tm>
- Fellows, I. (2018). *wordcloud: Word Clouds*. R package (version 2.6) [Software]. <https://CRAN.R-project.org/package=wordcloud>

- Friedman, J., Hastie, T. y Tibshirani, R. (2019). *glasso: Graphical Lasso: Estimation of Gaussian Graphical Models*. R package (version 1.11) [Software]. <https://CRAN.R-project.org/package=glasso>
- Gaikwad, S. V., Chaugule, A. y Patil, P. (2013). Text Mining Methods and Techniques. *International Journal of Computer Applications*, 85(17), 42-45. <https://doi.org/10.5120/14937-3507>
- Gentleman, R., Whalen, E., Huber, W. y Falcon, S. (2022). *graph: A package to handle graph data structures*. R package version 1.74.0.
- Gil Pascual, J. A. (2021). *Minería de texto con R: aplicaciones y técnicas estadísticas de apoyo*. UNED - Universidad Nacional de Educación a Distancia.
- Hennig, C. (2020). *fpc: Flexible Procedures for Clustering*. R package (version 2.2-9) [Software]. <https://CRAN.R-project.org/package=fpc>
- Kaneko, T., Asano, T. y Miwa, H. (2021). Estimating Ideal Points of Newspapers from Editorial Texts. *The International Journal of Press/Politics*, 26(3), 719–742. <https://doi.org/10.1177/1940161220935058>
- Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(1). <https://doi.org/10.1016/j.jjime.2021.100008>
- Lanzarini, L. C., Hasperué, W., Villa Monte, A., Jimbo Santana, P., Reyes Zambrano, G., Corvi, J. P., Fernandez Bariviera, A. y Olivas Varela, J. Á. (2019). *Minería de datos, minería de textos y Big Data* [Presentación de la conferencia]. XXI Workshop de Investigadores en Ciencias de la Computación, WICC 2019, Universidad Nacional de San Juan, Argentina. <http://sedici.unlp.edu.ar/handle/10915/77022>
- Lang, D. y Chien, G. (2018). *wordcloud2: Create Word Cloud by 'htmlwidget'*. R package (version 0.2.1) [Software]. <https://CRAN.R-project.org/package=wordcloud2>
- Linares Rodriguez, V., Paniagua Santamaria, P. y Abejón Mendoza, P. (2014). La matriz de la opinión en los periódicos: el artículo editorial. *Sphera pública*, 1(14), 142-173. <https://sphera.ucam.edu/index.php/sphera-01/article/view/171>
- López Arroyo, A. (2016). *El artículo editorial como texto argumentativo. Un estudio del género periodístico más razonador a la luz de la teoría de la argumentación* [Tesis doctoral, Universidad Complutense de Madrid]. Repositorio Institucional de la UCM. <https://eprints.ucm.es/id/eprint/41595/1/T38579.pdf>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. y Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.4 - For new features, see the 'Changelog' file (in the package source) [Software]. <https://CRAN.R-project.org/package=cluster>
- Moreno Espinosa, P. (2003). Géneros para la persuasión en prensa: los editoriales del Diario El País. Ámbitos: Revista Internacional de Comunicación, 9 y 10, 225-238. <http://hdl.handle.net/11441/67480>
- Moscovici, S. (1976). *Social influence and social change*. Academic Press.
- Moscovici, S. (2001). *Social representations: explorations in social psychology*. New York University Press.
- Moscovici, S. (2008). *Psychoanalysis: its image and its public*. Polity.
- Murrell, P. (2022). *grid Graphics*. <https://stat.ethz.ch/R-manual/R-devel/library/grid/doc/grid.pdf>
- Nava, M. y Marques, F. P. J. (2019). From “Leftist” To “President”: Journalism and Editorial Coverage of Brazil's Lula in Five Elections. *Journalism Practice*, 13(10), 1200-1221. <https://doi.org/10.1080/17512786.2019.1587640>
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>
- Núñez Ladevéze, L. y Vázquez Barrio, T. (2008). La información, la opinión y la editorial periodística. *Textual & Visual Media*, 1, 255-274. <https://textualvisualmedia.com/index.php/txtvmedia/article/view/22>
- Núñez Ladevéze, L. (2007). El contexto social de la empresa periodística en una sociedad libre. *Contexto: Revista de Comunicación de la Universidad de La laguna*, 1, 31-48.

- Pollak, S., Coesemans, R., Daelemans, W. y Lavrač, N. (2011). Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining. *Pragmatics*, 21(4), 647-683. <https://doi.org/10.1075/prag.21.4.07pol>
- Quispe Guillen, S. W. (2019). *Tratamiento periodístico de la vacancia de Pedro Pablo Kuczynski en los diarios El Comercio y La República, marzo de 2018* [Tesis de Licenciatura, Universidad César Vallejo, Lima, Perú]. Repositorio Institucional UCV. <https://repositorio.ucv.edu.pe/handle/20.500.12692/45379>
- Sinoara, R. A., Maracini, R. M. y Rezende, S. O. (2021). Mineração de Textos e Semântica: desafios, abordagens e aplicações. *Revista de Sistemas de Informação da FSMA*, (27), 41-53. Repositorio de la Universidad de São Paulo. <https://repositorio.usp.br/item/003038161>
- Suárez Paredes, A. J. (2013). *Análisis discursivo de la Central del diario Correo en las elecciones presidenciales 2011* [Tesis de Licenciatura, Pontificia Universidad Católica del Perú, Facultad de Ciencias y Artes de la Comunicación. Lima, Perú]. Repositorio Digital de Tesis y Trabajos de Investigación PUCP. <http://tesis.pucp.edu.pe/repositorio/handle/123456789/7564>
- Tineo Valencia, D. A. (2020). *Tratamiento Periodístico de la disolución del Congreso en las versiones digitales de Expreso y la República, Lima 2019* [Tesis de Licenciatura, Universidad César Vallejo. Lima, Perú]. Repositorio Institucional UCV. <https://hdl.handle.net/20.500.12692/50440>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind, New Series*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- Urbizagástegui-Alvarado, R. (2021). La bibliometría brasileña: minería de textos. *Revista ACB: Biblioteconomia em Santa Catarina*, 26(1), 1-18. <https://revista.acb.org.br/racb/article/view/1768>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wijffels, J. (2022). textplot: Text Plots. R package version (0.2.1) [Software]. <https://CRAN.R-project.org/package=textplot>
- Yuan, H. (2016). Measuring media bias in China. *China Economic Review*, 38, 49-59. <https://doi.org/10.1016/j.chieco.2015.11.011>

ANEXO

Paquetes de R utilizados para el análisis de los editoriales de *La República*

Nombre	Descripción	Empleado
tm Text Mining Package	Marco para las aplicaciones de minería de texto dentro de R (Feinerer y Hornik, 2020)	Paquete básico para realizar la minería de texto
readtext Import and Handling for Plain and Formatted Text Files	Funciones para importar y manejar archivos de texto y archivos de texto formateados con metadatos adicionales, como '.csv', '.tab', '.json', '.xml', '.html', '.pdf', '.doc', '.docx', '.rtf', '.xls', '.xlsx' y otros. (Benoit y Obeng, 2021)	Lectura de los archivos de la carpeta Republica que contienen los editoriales
SnowballC Snowball Stemmers Based on the C 'libstemmer' UTF-8 Library	Una interfaz R para la biblioteca 'libstemmer' de C que implementa el algoritmo de derivación de palabras de Porter para colapsar palabras en una raíz común para ayudar a comparar el vocabulario. (Bouchet-Valat, 2020).	Remoción de los stopwords, es decir, palabras sin significado como los artículos, preposiciones, etc.
ggplot2 Create Elegant Data Visualisations Using the Grammar of Graphics	Un sistema para crear gráficos de forma 'declarativa', basado en "La gramática de los gráficos". Usted proporciona los datos, le dice a 'ggplot2' cómo asignar variables a la estética, qué primitivas gráficas utilizar, y él se encarga de los detalles (Wickham, 2016)	Visualización de las frecuencias de los términos identificados y las asociaciones o correlaciones entre los términos elegidos
graph A package to handle graph data structures	Paquete que implementa algunas capacidades simples de manejo de gráficos. (Gentleman et al., 2022)	Trazado de las asociaciones entre las palabras
grid The Grid Graphics Package	Sistema gráfico de bajo nivel que proporciona un gran control y flexibilidad en la apariencia y disposición de la salida gráfica. (Murrell, 2022)	
RColorBrewer ColorBrewer Palettes	Proporciona esquemas de color para los mapas (y otros gráficos). (Neuwirth, 2014)	Color en nube de palabras y otros gráficos
wordcloud Word Clouds	Funcionalidad para crear bonitas nubes de palabras, visualizar diferencias y similitudes entre documentos, y evitar el exceso de trazado en los gráficos de dispersión con texto. (Fellows, 2018)	Creación de nube de palabras
wordcloud2 Create Word Cloud by 'htmlwidget'	Una herramienta de visualización rápida para crear wordcloud usando 'wordcloud2.js'. (Lang y Chien, 2018).	
fpc Flexible Procedures for Clustering	Varios métodos para la agrupación y validación de clústeres. (Hennig, 2020)	Creación del dendrograma, validar y estimar las agrupaciones. Agrupación de clústeres en torno a los términos más frecuentes.
cluster "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	Métodos para el análisis de conglomerados. (Maechler et al., 2022)	
igraph Network Analysis and Visualization	Rutinas para gráficos simples y análisis de redes. (Csardi y Nepusz, 2006)	Trazado de las redes de las palabras más representativas utilizadas en los editoriales de <i>La República</i>
textplot Text Plots	Visualiza relaciones complejas en textos. Esto se hace proporcionando funcionalidades para mostrar redes de co-ocurrencia de texto, redes de correlación de texto, relaciones de dependencia, así como agrupación de texto e 'incrustaciones' de texto semántico (Wijffels, 2022)	Para el gráfico de barras de las frecuencias de las palabras
glasso Graphical Lasso: Estimation of Gaussian Graphical Models	Estimación de una matriz de covarianza inversa dispersa utilizando una penalización de Lasso (L1). Se proporcionan facilidades para estimaciones a lo largo de un camino de valores para el parámetro de regularización. (Friedman et al., 2019)	Creación del gráfico de la correlación de las palabras (redes de palabras)