Rev. Elet. Gestão e Serviços
V.13, n.1, Jan./Jun. 2022

ARTIGO ORIGINAL
ORIGINAL ARTICLE

ANÁLISIS COMPARATIVO DEL SEGUIMIENTO AL HASHTAG #DÍASINIVA EN COLOMBIA A TRAVÉS DE TWITTER DURANTE LA PANDEMIA DE COVID-19

COMPARATIVE ANALYSIS OF FOLLOWING THE HASHTAG #DÍASINIVA IN COLOMBIA THROUGH TWITTER DURING THE COVID-19 PANDEMIC

Katherin Pérez Mendoza[1]
Carlos Busón[2]
Mónica Pineda Arroyo[3]

Resumen

El presente trabajo trata sobre la evidencia de los posibles resultados del uso de la textometría como herramienta para el análisis de representaciones e interacciones sociales en una red social sobre hechos concretos. Al mismo tiempo, el ejercicio realizado sirve como ejemplo para comprender con precisión las reacciones de una población en torno a una política económica y de salud pública, teniendo en cuenta que el análisis realizado fue sobre datos recopilados de Twitter a través del hashtag #diasiniva sobre un serie de medidas tributarias tomadas por el gobierno colombiano en la primera fase de la pandemia de Covid-19 en 2020. Esta investigación es de naturaleza cualitativa y busca utilizar el software Iramuteq para comprender el sentimiento general de la población colombiana hacia una determinada política en un momento dado. tiempo específico de crisis. La herramienta presenta gráficos que permiten un análisis sencillo de grandes volúmenes de presente información

Palabras clave: Análisis cualitativo; Colombia; impuestos; COVID-19; Twitter

Abstract

The present work deals with the evidence of two possible results of the use of textometry as a tool for the analysis of social representations and interactions in a social network on specific facts. At the same time, the exercise carried out serves as an example to understand precisely the reactions of a population around an economic and public health policy, bearing in mind that the analysis carried out was on data collected from Twitter through the hashtag #diasiniva on A series of tax measures taken by the Colombian government in the first phase of the Covid-19 pandemic in 2020. This research is of a qualitative nature, seeking to use the Iramuteq software to understand the general sentiment of the Colombian population in relation to a certain policy at a time crisis specific. The tool presents graphs that allow a simple analysis of large volumes of information.

Key Words: Qualitative analysis; Colombia; taxation; Covid-19; Twitter

[1]Investigadora y Directora de Programa de Economía de la Universidad de La Guajira. Ha realizado investigaciones en valoración económica de ecosistemas desde la diversidad cultural utilizando herramientas tecnológicas novedosas como estrategia de ordenamiento territorial y desarrollo local. Universidad de La Guajira UNIGUAJIRA - Colombia.E-mail: kperezm@uniguajira.edu.com

[2] Investigador en la Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do en colaboración con el Conselho Nacional de Desenvolvimento Científico e Tecnológico (FUNDECT/CNPq).. E-mail: cbuson@gmail.com

[3] Universidad de La Guajira UNIGUAJIRA - Colombia.E-mail: mcpineda@uniguajira.edu.com

Artigo recebido em: 09 de março de 2022. Artigo aceito em 31 de julho de 2022.

1 INTRODUCCIÓN

La COVID-19, esa enfermedad que empezó como un virus de preocupación que medio mundo seguía como algo lejano desde las noticias y comunicados de la Organización Mundial de la Salud en sus primeras semanas para finales de 2019, se ha convertido en el eje que ha modificado casi todas las estructuras sociales y económicas durante más de un año. Ya para la segunda quincena de marzo de 2020 los problemas derivados de la pandemia de COVID-19 provocaron un cierre masivo de todas las fronteras internacionales y, lo que había empezado como un problema local en la ciudad de Wuhan en la República Popular China, se extendió rápidamente por todo el globo. Era una enfermedad desconocida, pero con ratios de contagio alarmantes y porcentajes de letalidad elevada, características que presionan y preocupan a cualquier sistema de salud existente.

Así las cosas, las autoridades sanitarias y gobiernos se vieron obligados a medidas drásticas de aislamiento social y confinamiento, con el fin de detener el avance de lo que se convirtió en pocas semanas en una pandemia. Lo anterior provocó un auténtico terremoto en las economías de los países afectados: el cierre de fronteras, la parada forzada de diversas cadenas productivas por culpa del obligado aislamiento social, la caída de las bolsas de valores internacionales, la incertidumbre de los inversionistas, entre otras muchas más cosas, provocaron una situación sin precedentes en la sociedad y economía globales. Cada nación intentó sortear sus efectos como pudo, de forma aislada, y no coordinada con sus vecinos, lo que sin duda agravó los efectos de la misma.

Ante lo acontecido, la economía de los países se vio gravemente afectada, y los problemas que venían ralentizando el crecimiento económico de múltiples países desde la crisis económica de 2008 y 2009, se hicieron sentir con mayor voracidad. Según datos de la Comisión Económica para América Latina y el Caribe (CEPAL, 2020, p. 8), el mayor impacto económico (así como en estadísticas sanitarias) se registra en los países de Latinoamérica, lo que arroja una proyección de la contracción del PIB regional de aproximadamente 9,1% para el año 2020, cifra que se refleja en "(…) el cierre de 2,7 millones de empresas, 44,1 millones de desocupados, 231 millones de personas en condición de pobreza, de las cuales 96 millones se encontrarán en situación de pobreza extrema, y aumentos significativos del índice de Gini". (GINI, 2020)

Es en este desafortunado escenario en el que se encuentra inmerso Colombia, pues

según el Departamento Administrativo Nacional de Estadística (DANE), el PIB nacional sufrió una contracción aproximada del 6,8% para el año 2020, con una tasa de desempleo promedio de 16,1%[1].

Lo anterior preocupa aún más si se tiene en cuenta la alta vulnerabilidad de los empleados ante la proporción de informalidad que en promedio alcanzó el 47% de los empleos nacionales para el mismo periodo.

Dicho esto, a pesar de que en el país establecieron medidas estrictas de aislamiento preventivo obligatorio desde el 26 de marzo de 2020[2] por un periodo inicial de 20 días, éstas fueron extendidas y flexibilizadas gradualmente para permitir la operación de cada vez más sectores económicos en medio de lo que se estableció como nueva normalidad.

Lo anterior contrasta con las cifras nacionales referentes a la pandemia, pues cinco meses después del inicio de la emergencia económica, social y ecológica[3], el país atravesaba el primer pico fuerte desde la llegada de la pandemia al territorio, y el avance del virus lo ubicaba como uno de los focos más grandes de contagio de la región con aproximadamente 541.000 casos positivos y más de 17.316 muertes, según los registros de la OMS[4] en agosto de 2020, lamentablemente dichas cifras en agosto de 2021 estos datos superan los 4.867.000 casos positivos y los 123.000 fallecidos.

Los datos muestran entonces que, a pesar de que Colombia pudo controlar el avance de la pandemia durante los primeros meses de la misma, a diferencia de otros países de la región como Brasil (donde ha impactado de manera drástica en la sociedad por evitar las medidas de aislamiento social más estrictas y ha mantenido el desafortunado liderato latinoamericano respecto a casos y muertes), no logró escapar del flagelo en términos de salud pública ni económicos.

Las disyuntivas entre la crisis económica y la presión sobre el aparato de salud nacional se convirtió en el centro del análisis desde los primeros días del contexto, teniendo en cuenta que los pronósticos del Fondo Monetario Internacional (FMI) (y en general de todas las entidades referentes en la materia) de crecimiento económico para Colombia se

[1] Promedio simple tomado de los datos suministrados por el DANE derivados de la Gran Encuesta Integrada de Hogares respecto a las series desestacionalizadas de la proporción de informalidad según tamaño de empresa hasta 5 trabajadores para los meses comprendidos entre mayo y diciembre de 2020 (Para los meses de enero, febrero, marzo y abril la entidad no registra información).

[2] Decreto 457 de 2020, el cual se imparten instrucciones en virtud de la emergencia sanitaria generada por la pandemia del Coronavirus COVID-19 y el mantenimiento del orden público.

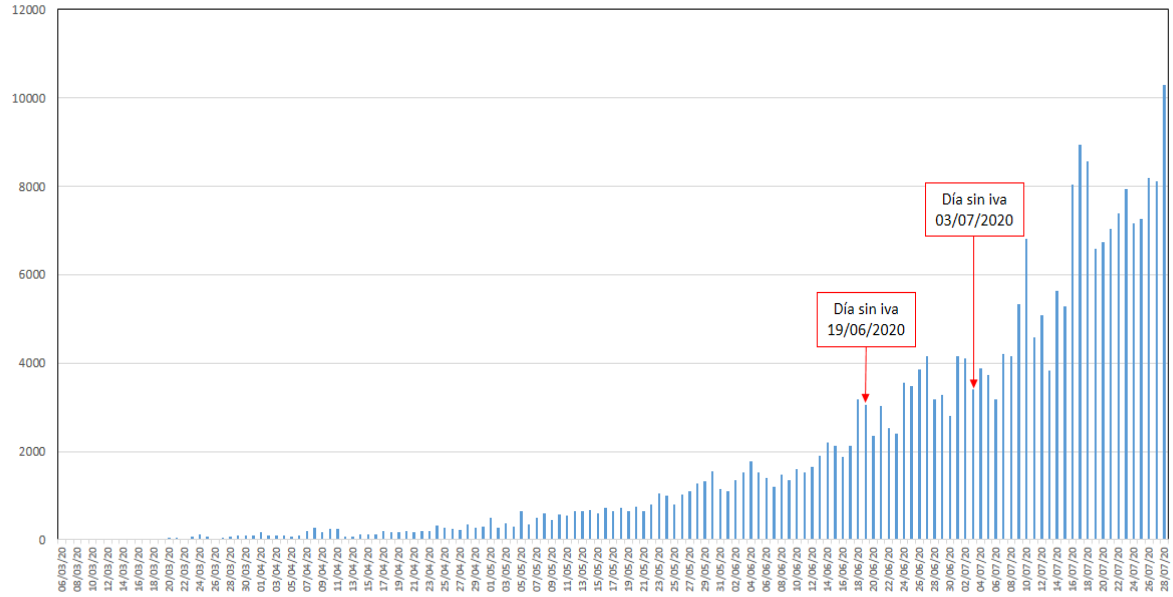
[3] Decreto 417 de 2020, por el cual se declara Estado de Emergencia Económica, Social y Ecológica en todo el territorio Nacional.

[4] Datos tomados del WHO Coronavirus (COVID-19) Dashboard, a corte del 25 de agosto de 2020. Y 10 de agosto de 2021. Recuperado de: https://covid19.who.int/table.

convirtieron en porcentajes negativos de -2,4% para febrero de 2020, y fueron empeorando con cada prolongación de las medidas tomadas para contener la pandemia, llegando incluso al orden de -8,2% para octubre del mismo año.

En aras de estimular la recuperación económica del país en medio de todo lo expuesto anteriormente, el gobierno colombiano mediante Decreto 682 de 2020 estableció la exención especial del impuesto sobre las ventas (IVA) en el marco de una política económica que, inicialmente, oficializaba tres días no consecutivos de liberación de dicho impuesto a las compras realizadas durante las 24 horas de esas fechas específicas para un conjunto de bienes determinados y que fueran adquiridos por personas naturales como consumidores finales.

Figura 1. Línea de tiempo entre el 06/03/2020 al 28/07/2020 de los contagiados por el covid-19 en Colombia



Fuente: datos recuperados el 29/07/2020
<http://datosabiertos.esri.co/datasets/782122624f364fbdbd7e287b96c4a358_6/data> y elaboración propia.

Ante la llegada del primer día sin IVA el 19 de junio de 2020, se presentaron aglomeraciones e infracciones de las medidas sanitarias establecidas para garantizar la prevención y mitigación del virus justo para los días que se registrarían, semanas más tarde, como el inicio del primer pico de la pandemia en Colombia. No obstante, hubo un parte esperanzador en cuanto a reactivación económica, en tanto que la Dirección de Impuestos y Aduanas Nacionales DIAN, así como la Federación Nacional de Comerciantes FENALCO, dieron un reporte positivo que destacó la cifra de 5 billones de pesos en venta en comercio

durante la jornada.

Teniendo en cuenta los resultados poco atractivos desde la salud pública, las ciudades capitales del país tomaron medidas más restrictivas para el segundo día sin IVA, llevado a cabo el 03 de julio de 2020, por lo que las transacciones se realizaron mayoritariamente de manera virtual y las presenciales se hicieron bajo mayores controles sanitarios.

Pese a los ajustes mencionados para el segundo día sin IVA, el aumento escalonado de las cifras de contagios y muertes por COVID-19 ya era más que evidente para la primera mitad de julio, así como el alcance del límite en las ocupaciones de Unidades de Cuidado Intensivo en los hospitales de las principales ciudades del país, lo que generó el aplazamiento del tercer día (que sería inicialmente el 19 de julio) hasta el 21 de noviembre de 2020 mediante el Decreto 1474 de 2020.

Como toda medida o política pública, y más aquellas que contraponen dos aspectos de suma relevancia e interés colectivo cómo la salud pública y la economía, la política de días sin IVA fue presentada como un salvavidas al sector más golpeado de la economía, pero generó controversias de opinión que se expresaron explícitamente en un debate de la ciudadanía, que se apropió de las redes sociales como escenario propicio para su desarrollo. En este sentido, en la plataforma de la red social Twitter, los usuarios utilizaron el hashtag #díasininva para relacionar los comentarios vinculados al tema y mantener un hilo de expresiones que llegaron a ser tendencias nacionales.

2 REVISIÓN DE LA DOCUMENTACIÓN

La minería de datos de redes sociales es una importante herramienta que nos permite acceder a grandes volúmenes de información para su posterior tratamiento. Precisamente, el trabajo de Bouchet e Bastin (2013) sobre la minería de datos en Internet nos muestra algunas de las numerosas posibilidades de análisis. Así mismo, existen diversas aplicaciones de los mismos, como los análisis de sentimientos en Twitter, citados por Baviera (2017). En nuestro trabajo, en esta primera fase, buscamos aplicar una metodología diferente y analizar los tweets mediante la textometría.

Trabajar con textos mediante estas tecnologías son una interesante opción para analizar comentarios de las redes sociales, en los trabajos de Dijk (2000) en "*El estudio del discurso. En El discurso como estructura y proceso"* son extremadamente útiles para

trabajar esta línea de investigación. Dijk (2002), nos habla del "*uso actual de la lengua por usuarios concretos en situaciones sociales concretas*." Esta es una interesante opción para aplicar el Análisis Crítico del Discurso o CDA, pues plantea el uso interdisciplinar del estudio del discurso, teniendo en cuenta cualquier manifestación lingüística, como sería el caso del estudio de los tweets. Dicha metodología permite observar las relaciones sociales frente a un evento concreto y el uso del lenguaje que se realiza.

Por otro lado, las investigaciones desarrolladas por Arrieta e Avendãno (2018) sobre el análisis lingüístico, sociodiscursivo y sociopragmático en los discursos de los tweets, nos sirven para valorar esa opción y analizar una muestra representativa con el fin de obtener pautas de comportamiento frente a un evento.

Los estudios sobre los comentarios nos dicen mucho sobre el pensamiento colectivo de un grupo social, nos revelan sus pautas, pues según Busón (2020)

> Siempre hay una fuente, una audiencia un contenido con una intención que busca provocar un efecto: "podemos aprender mucho sobre nosotros mismos y las formas en que otras personas tratan de explotar el valor de nuestro ser social". La falta de pensamiento crítico en la educación, cada vez más despreciada y atacada por gobiernos de talante autoritario, lleva a los usuarios a no cuestionarse lo que realmente se esconde tras los comentarios de particulares, de grupos de opinión o de empresas, que buscan modificar pautas a través de la polémica. Este es un conformismo silencioso que es necesario estudiar. (BUSON, 2020, p. 121)

En síntesis, el análisis de datos en Twitter nos permite observar las pautas del comportamiento de los usuarios frente a un evento concreto. Las respuestas permiten infinitud de estudios desde los análisis de sentimientos, así como geolocalización de los tweets para elaborar mapas de calor. En este trabajo quisimos estudiar la respuesta ante una situación concreta como lo es la iniciativa del día sin IVA en Colombia, utilizando técnicas de análisis semántico basado en R y que funcionan a través de IRaMuTeQ.

A través de los algoritmos de esta herramienta pudimos analizar dos muestras de datos para reflexionar sobre el sentir general ante un evento, los tweets generados en cada una de las dos primeras fechas del día sin IVA en Colombia en medio de la pandemia por COVID-19. Diversos trabajos previos como los de Ratinaud (2014). Ratinaud(2016). Sarrica, Mingo, Mazzara e Leone (2016), Baviera (2017) nos muestran los recursos disponibles y la utilidad del uso de IRaMuTeQ en análisis textual para tratar datos cualitativos.

Según Longhi e Saigh (2016 apud PINCEMIN, 2011), la textometría[5] ofrece un

[5] Traducción libre "La textometría, nacida en Francia en los años 80, ha desarrollado técnicas poderosas para

enfoque instrumentado para el análisis de un corpus de texto, además de proporcionar procedimientos de clasificación y cálculos estadísticos para el estudio de corpus digitales de los mismos. Además, establece modelos contextuales y contrastantes. Así, el texto se caracteriza por sus palabras en relación con su uso en el corpus, la palabra se caracteriza por sus coincidencias, entre otras opciones de gran utilidad.

Este método es particularmente relevante para la exploración del corpus en investigaciones cualitativas como nos muestran los trabajos de Justo e Camargo (2014), Souza, Thuler, Lowen e Peres (2018), donde desde 1980, con el advenimiento de la computación, se produjo un aumento significativo de herramientas informáticas denominadas CAQDAS (Computer Aided Qualitative Data Analysis Software). Éstas últimas permitieron abrir nuevos campos de la investigación aplicada en ciencias humanas y sociales, no obstante, pueden ser empleadas con bastante eficiencia en otras áreas, como es el caso del estudio de pacientes clínicos, por ejemplo.

Simultáneamente, nos permite una observación detallada y global de diferentes textos mientras permanece cerca de ellos, y destaca el hecho de que el lenguaje es un campo de observación importante para los humanos y las ciencias sociales.

Entre los diversos instrumentos existentes, nos decantamos por utilizar una herramienta basada en software libre el IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), al ser un paquete de uso relativamente sencillo que utiliza R como motor para los cálculos estadísticos. Desarrollado por Pierre Ratinaud (1993), es un uso relativamente sencillo contando con varias traducciones y diccionarios para diversas lenguas entre ellas el español.

Las posibilidades que ofrece IRaMuteQ en el tratamiento de muestras de datos desde entrevistas, encuestas, textos, y los comentarios que quisimos experimentar, son una interesante opción de cara a procesar datos cualitativos en diferentes estadísticas de textos, Por lo tanto, ante el uso de esta herramienta se cuenta con interesantes posibilidades de analizar discursos o los comentarios de las redes sociales como una forma de establecer pautas de comportamiento de los usuarios frente a un evento concreto.

Dicha tecnología ya es utilizada en estudios de mercado para seguimiento del

el análisis de gran cantidad de textos. Después de la lexicometría y el análisis estadístico de texto, ofrece herramientas y métodos probados en múltiples ramas de las humanidades y está estadísticamente bien fundado. El proyecto reúne los desarrollos de software de Textometría de código abierto para configurar una plataforma modular llamada TXM. Es a la vez una herencia de influencia internacional y el lanzamiento de una nueva generación de investigación textométrica, en sinergia con las tecnologías de corpus existentes (Unicode, XML, TEI, herramientas de PNL, CQP, R)" .http://textometrie.ens-lyon.fr/?lang=en

impacto de marca, no obstante, en las investigaciones académicas aún sigue siendo poco utilizada. Así las cosas, con este trabajo pretendemos profundizar en esta interesante línea de investigación y contar con las diversas utilidades antes mencionadas en el estudio de los fenómenos sociales.

3 PROCEDIMIENTOS METODOLÓGICOS

Fueron obtenidos 2050 comentarios alojados en la red social de microblogueo Twitter sobre el día sin IVA bajo el hashtag #diasiniva para el día 19 de junio de 2020. Así mismo, se obtuvieron 2080 comentarios para el día 03 de julio de 2020 en este mismo escenario, Lo anterior se realizó bajo la técnica de *web scraping*, a través de la cual pudimos obtener todos los comentarios en formato CSV, los cuales fueron posteriormente tratados en la hoja de cálculo de libre Office y el Notepad ++, con el fin de constituir el corpus de datos de dichas muestras las cuales, cabe mencionar, fueron trabajadas y analizadas de forma independiente en el software IRaMuTeQ (tabla 1).

Tabla 1. Datos de las muestras analizadas en IRaMuTeQ y R.

Datos	19 de junio de 2020	03 de julio de 2020
Número de textos	n=2.055	n=2.080
Número de segmentos de texto	2.056	2.131
Número de formas	3.764	4.295
Número de ocurrencias	49.015	71.019
Número de hápax	1.286 (2,59%de ocurrencias - 46.36% de formas)	1.459 (2,05%de ocurrencias - 42.74% de formas)
Media de ocurrencias por texto	24,15	34,14
Número de lemas	2.773	3.414
Número de formas activas	2.539	3.142
Número de formas suplementarias	153	48
Número de formas activas con frecuencia	>= 3: 1018	>= 3: 1275
Media de formas por segmento	23,839981	33,326607
Número de clases	3	3
Clasificación textos	1.629 textos clasificados en 2.055 (79,27%)	1.482 textos clasificados en 2.080 (71,25%)

Fuente: Tweets extraídos en vicinitas.io y elaboración propia en IRaMuTeQ y R.

En total analizamos dos muestras para tratar los datos del #díasiniva, de los días 19 de junio de 2020 (n=2055) y del 03 de julio de 2020 (n=2080). El resultado de todo este proceso de tratamiento de la información formó los dos corpus textuales para su posterior análisis en IRaMuTeQ.

Dichos corpus textuales incluyen formas activas, es decir, substantivos, verbos

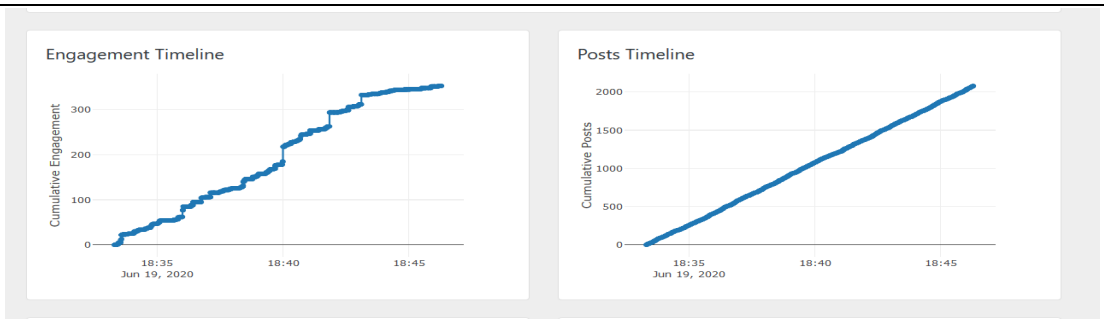
(suplementarias o no), adjetivos (suplementarias o no), adverbios (suplementarias o no) y formas no reconocidas y las formas suplementarias que, en estos casos, incluyen los substantivos suplementares y adjetivos numéricos. En este estudio, las formas activas fueran denominadas palabras activas.

En la figura 2, podemos observar las líneas de tiempo de los materiales obtenidos mediante la técnica de *web scraping* en *vicinitas.io*. De acuerdo al tiempo de extracción de los Tweets durante los dos días podemos observar que la interacción en el primer día (19 de junio de 2020) fue muchísimo mayor teniendo en cuenta que se obtuvieron 2055 tweets en un término de 13 minutos, mientras el segundo día (03 de julio de 2020) los tweets fueron 2080 en el lapso de tiempo de 2 horas y media.

La muestra es un pequeño pero significativo fragmento de los tweets que nos dan una imagen del sentir sobre el *hashtag* #díasiniva. De allí pudimos con ello valorar aspectos sociales frente a la medida económica.

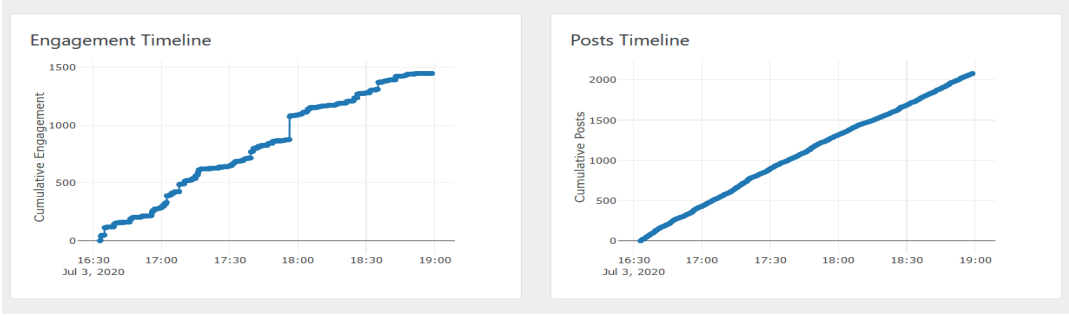
En la figura 3 podemos identificar la distribución de los mensajes que fueron distribuidos a partir de las categorías que permite la aplicación Twitter, por tipo de post (*tweets*, *retweets* y *replies* o respuestas) y por contenido media (fotos, videos y GIF's). Por lo tanto, se observa que los *retweets* fueron la pauta con un 82, 5% en los dos días, mientras las respuestas fueron de 2,45% en el primer día y 1,49% en el segundo. En cuanto a los tweets, que son los textos iniciales de los que derivan las interacciones anteriores, la cifra en el primer y segundo día fue de 15,1% y 16% respectivamente.

Figura 2. Comparación de las líneas de tiempo de los posts de las muestras analizadas en este estudio[6].



Línea de tiempo de los 2055 tweets extraídos el 19 de junio de 2020 comprendidos entre 8:33 y las 18:46

[6] https://www.vicinitas.io/free-tools/download-search-tweets?tracker=%23D%C3%ADasiniva



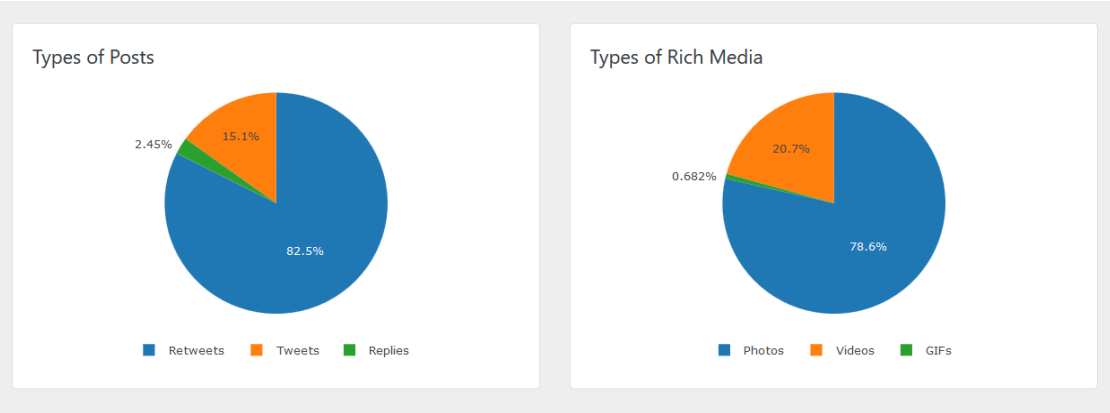
Línea de tiempo de los 2080 tweets extraídos el día 03 de julio de 2020 comprendidos entre 16:30 y las 19:00

Fuente: Tweets extraídos en vicinitas.io y elaboración propia.

Así mismo, en cuanto al contenido media, la pauta fueron las fotos y videos, con un 78,6% y 20,7%, respectivamente, durante el primer día; mientras el segundo día las fotos fueron un 90% y los videos un 9,55%, dejando rezagados a los GIF's pues ocuparon un 0.682% y 0.482% del contenido utilizado durante los dos días.

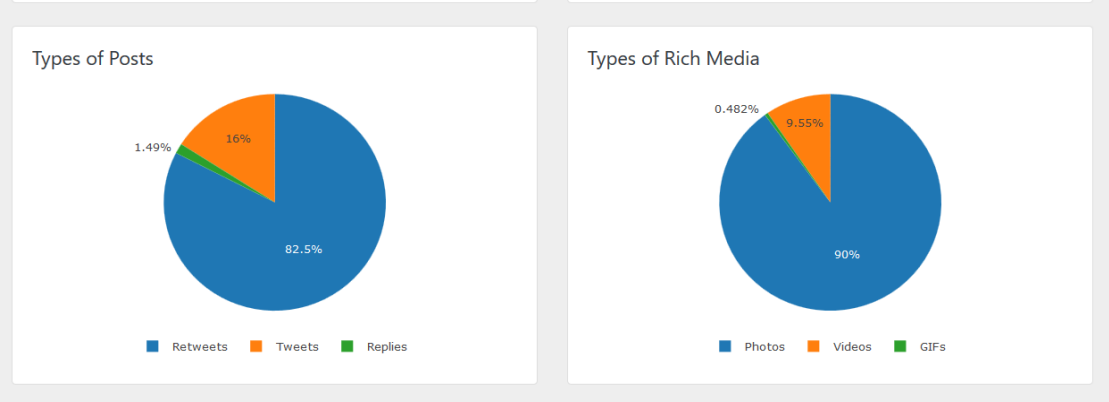
Aunque aún no se ha entrado en detalle al análisis de los datos, la anterior distribución en un primer indicio de cómo los individuos utilizan las redes sociales, y en especial Twitter como escenario de encuentro ideal para expresión de ideas bajo el formato de mensajes cortos y microblogueo, en torno a la interacción de las representaciones individuales y colectivas respecto a un evento concreto: El hecho de hacer retweet implica en gran medida la aceptación del mensaje inicial, las respuestas giran en torno al enriquecimiento de la discusión o debate frente a la temática, así como el gran porcentaje de videos e imágenes buscan mostrar la realidad bajo dichas representaciones visuales.

Figura 3. Comparación de los tipos de posts de las muestras analizadas en este estudio[7].



[7] https://www.vicinitas.io/free-tools/download-search-tweets?tracker=%23D%C3%ADasiniva

Tipo de posts empleados y sus contenidos de los 2080 tweets extraídos el día 19 de junio de 2020 comprendidos entre 18:33 y las 18:46



Tipo de posts empleados y sus contenidos de los 2080 tweets extraídos el día 03 de julio de 2020 comprendidos entre 16:30 y las 19:00

Fuente: Tweets extraídos en vicinitas.io y elaboración propia.

Para analizar los textos de los discursos se utilizó el método de nube de palabras, que agrupa las palabras únicamente en función de la frecuencia con que aparecen en el texto. Es un método sencillo, pero de utilidad a la hora de representar gráficamente el peso de ciertas palabras en el contenido, lo que permite con un simple vistazo apreciar la importancia de las mismas en el texto.

Asimismo, realizamos los análisis de similitud. Está es una técnica fundamentada en la teoría de los grafos, el cual permite la identificación de las ocurrencias de las palabras en un texto, donde el resultado gráfico son las indicaciones de conectividad entre las mismas. Lo anterior, permite identificar las ocurrencias de las palabras, su importancia y la forma en que estas se conectan.

De manera complementaria, se realizó un Análisis Jerárquico Descendiente (CDH), propuesto por el método de Reinert en que los segmentos de texto (fragmentos de los textos) son clasificados de acuerdo con sus respectivos vocabularios y sus conjuntos repartidos en formas más reducidas, así como el Análisis Factorial de Correspondencia (AFC), que se hace a partir de los datos obtenidos en la CHD de cada corpus textual, el cual permite un gráfico con ejes propuestos por el modelo de Reinert, en que los segmentos de texto son clasificados de acuerdo con sus respectivos vocabularios y sus conjuntos distribuidos en formas más reducidas.

4 DISCUSIÓN Y ANÁLISIS DE LOS DATOS

A continuación, se describen entonces los resultados del análisis de los discursos y comentarios de los tweets que utilizaron el hashtag #díasiniva mediante el software, dividiéndolos en las dos muestras representativas tomadas, teniendo en cuenta que cada una se generó en escenarios económicos, sanitarios y sociales distintos a pesar de la cercanía de los días en términos de días calendario.

Análisis de los tweets del #diasiniva del 19 de junio de 2020

Veamos ahora algunos de los resultados obtenidos. Para efectos de mejorar el análisis de las palabras relevantes alrededor de la etiqueta *díasiniva* decidimos eliminar esta de la nube que muestra la Figura 4, aunque esta fue la más citada por claras razones ((1846 veces). Aclaramos que subyacente a la nube, la etiqueta permanece presente en este y en los siguientes análisis realizados.

Figura 4. Nube de palabras de los tweets analizados del 19 de junio de 2020.



Fuente: Twitter y elaboración propia en IRaMuTeQ y R.

Encontramos entre las 10 primeras palabras que más se repiten en los discursos analizados las siguientes: *comprar* (323 veces), *hacer* (321 veces), *duque* (315 veces), *medida* (268 veces), *decir* (261 veces), *gente* (260 veces) e *inteligente* (257 veces). Asimismo, se observan palabras alrededor como *aglomeración, culpar, vida, bioseguridad, Colombia, gobierno*, economía, etc.

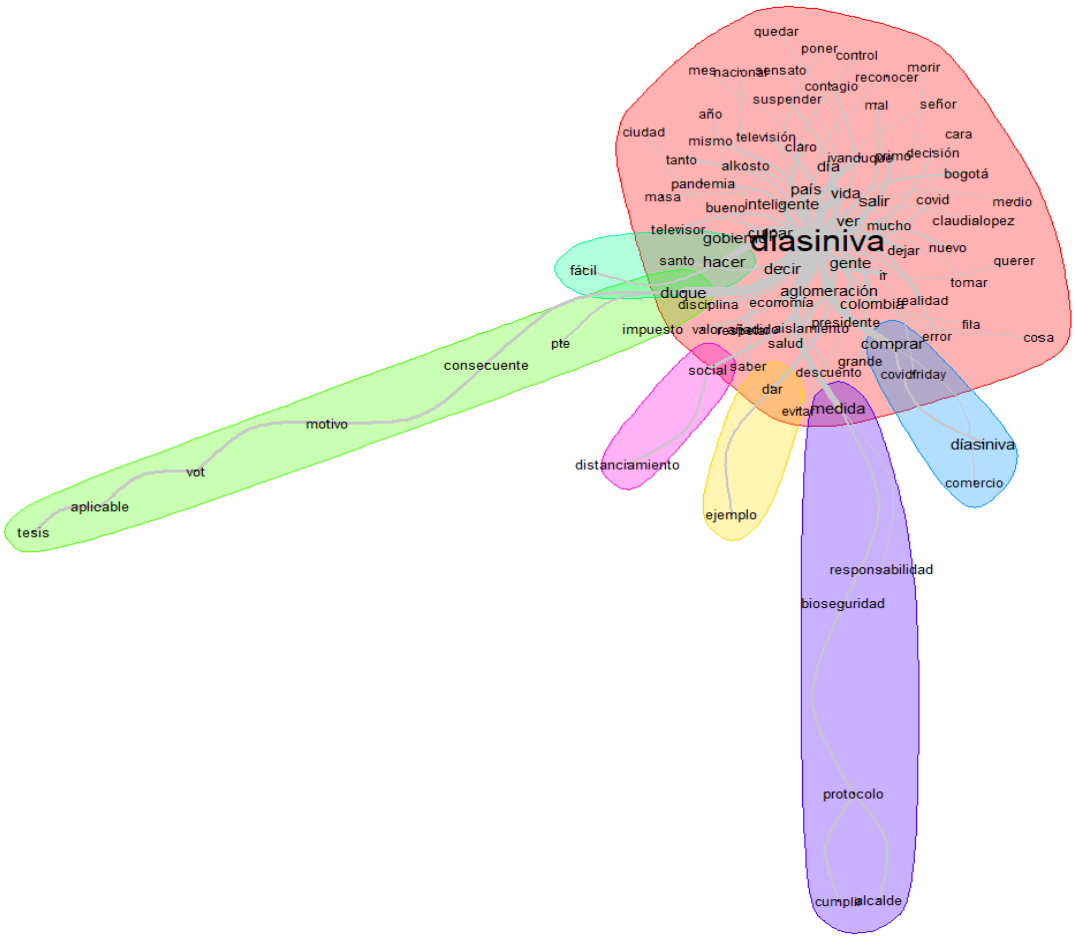
El análisis de similitud realizado (figura 5) con los textos analizados en el corpus nos permite encontrar un gran grupo de palabras conectadas entre sí: *Diasiniva* del que derivan otros grupos palabras conectadas. Algunos tienen que ver con la bioseguridad, el comercio, el distanciamiento social, etc. Lo hemos representado con los halos para destacar

los grupos de palabras conectadas y la relación que existe entre ellas y las diversas dependencias de las palabras que existen entre sí.

Tanto la nube de palabras como el análisis de similitud son conducentes a mostrar la disyuntiva presente en este primer día sin IVA en torno a la preponderancia de la economía y la salud pública, en tanto que los discursos de los usuarios giran alrededor de estos dos grandes tópicos.

Este es el punto en el que nos permitimos recordar el contexto en el que se desarrolla este día sin IVA y en el que se enmarcan los discursos encontrados: un país que viene de enormes restricciones sociales y cuarentenas obligatorias desde finales de marzo de 2020, medidas que venían repercutiendo enormemente en la economía nacional durante los casi tres meses que llevaban impuestas (aunque flexibilizadas desde su inicio) y en el que no se había declarado hasta el momento un primer pico de pandemia.

Figura 5. Análisis de similitud de los discursos analizados de la muestra del 19 de junio de 2020.



Fuente: Twitter y elaboración propia en IRaMuTeQ y R.

Revista Eletrônica Gestão e Serviços v.13, n. 1, pp. 3709 - 3730, Janeiro/Junho 2022.
ISSN Online: 2177-7284 e-mail: regs@metodista.br

Para generar un análisis de mayor amplitud, realizamos la Clasificación Jerárquica Descendente (CDH) (figura 6) en el corpus textual de los tweets del 19/06/2020, que formó un esquema jerárquico de clases de los vocabularios presentes en el corpus. Mediante el CDH pudimos obtener el contenido de cada una de las classes de los textos analizados de los discursos, en este caso fueron 3 (tres) *classes*; la *classe* 1 con un 77 % hace referencia temas como *Colombia, diasiniva, televisor, covid, impuesto valor añadido*. Dicha *clase* está relacionada con la *classe* 2 con un 12,1 % y destacan las palabras *ejemplo, bioseguridad, protocolo*. La *classe* 3 con un 10,9% hace referencia a las palabras entre otras a: *inteligente, duque, vida, aislamiento, economía, etc.*

Figura 6. Dendograma de las tres clases lexicales obtenidas de la clasificación jerárquica descendente de la muestra del 19 de junio de 2020.

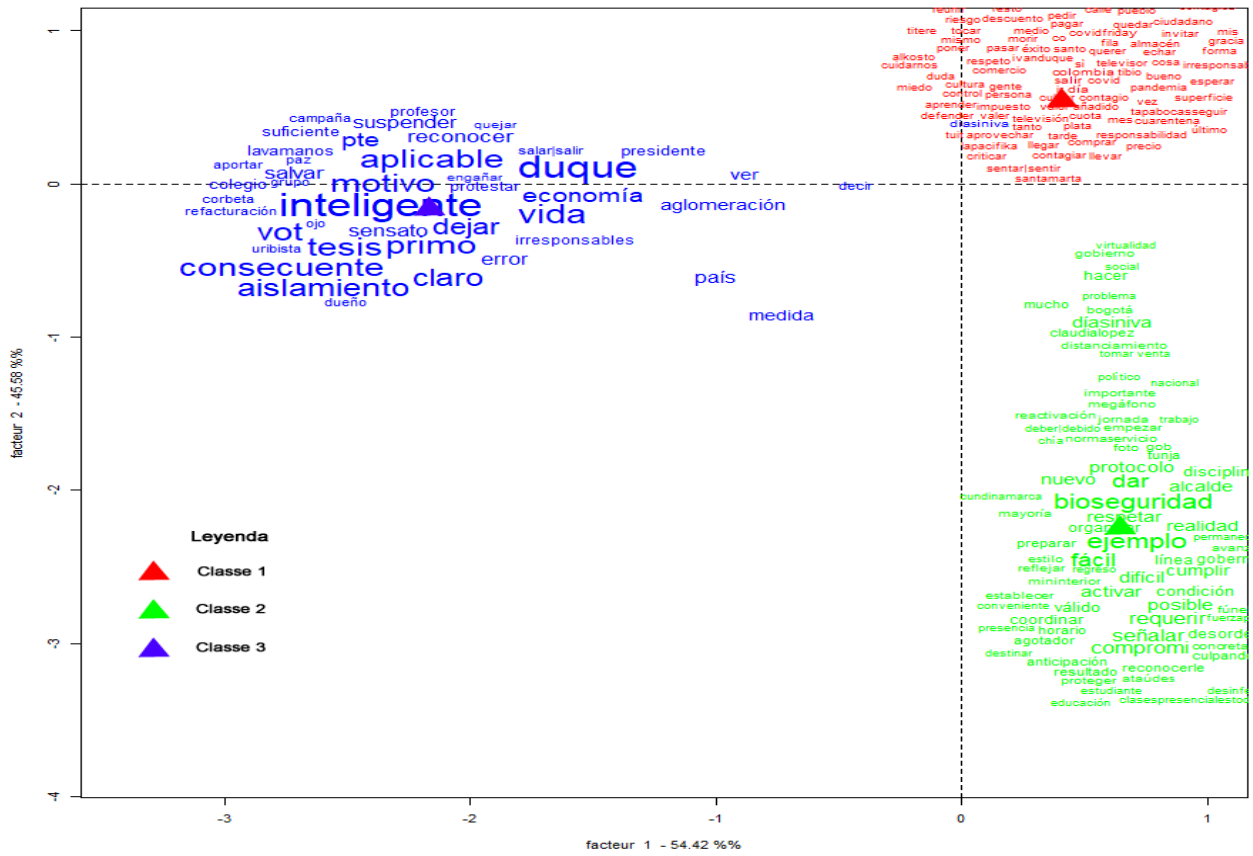


Fuente: Twitter y elaboración propia en IRaMuTeQ y R.

Por último, el análisis del AFC (figura 7) resultó en tres *classes*, en el eje X están representados en el factor 1 por el 54,42 % y el Y con un factor de 45,58 % del cuerpo textual. La distribución de clases quedó de la siguiente manera: la *classe* 1 quedó localizada en los ejes positivos de X e Y de forma que se encuentra en el cuadrante superior izquierdo, la *classe* 2 quedó restringida a la parte positiva de X no obstante, está localizada en la parte negativa del eje Y. Mientras que la *classe* 3 se encuentra en la parte negativa de X aunque distribuida de forma homogénea, tanto en la parte negativa, como la positiva del eje Y. Es posible observar que las palabras se representan muy bien definidas en los cuadrantes del plano cartesiano.

Así las cosas, podemos distinguir como los discursos en torno al #diasiniva expresaron principalmente expectativas y opiniones a priori de lo que la medida en sí significaría para la economía y salud nacional, además de mencionar de manera principal la responsabilidad de los administradores públicos ante la toma de estas medidas, al expresar en medio de los textos un número importante de veces a la presidencia y las alcaldías de las ciudades principales.

Figura 7. Análisis de Especificidades y AFC de las tres classes lexicales de la muestra del 19 de junio de 2020.



Fuente Twitter y elaboración propia en IRaMuTeQ y R.

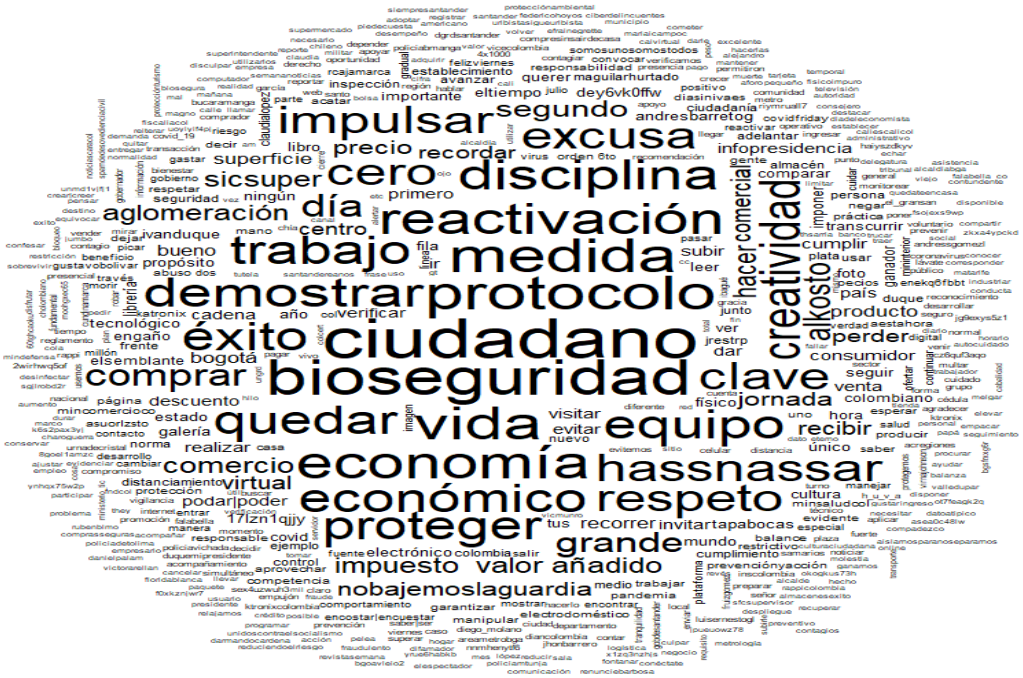
Los textos analizados son entonces, una muestra representativa del debate propio de una medida que ofrece una disyuntiva poco fácil de manejar entre medidas económicas y de salud pública, poniendo sobre la mesa las posibles consecuencias y responsabilidades de la misma, en términos incluso de análisis de costo beneficio.

Análisis de los tweets del #diasiniva del 03 de julio de 2020.

Para construir la nube de palabras en esta segunda fecha (figura 8), retiramos las dos palabras más citadas *diasiniva* (1122 veces), *días iniva* (964 veces)[8], de forma que pudiéramos observar con más detalle las otras palabras presentes en los tweets con mayor facilidad (sin que ello demerite el nivel de análisis a lograr).

Ante el escenario de la nube de palabras ilustrado en la figura 8, podemos encontrar entre las 10 primeras palabras que más se repiten en los discursos analizados fueron: *ciudadano* (641 veces), *bioseguridad* (575 veces), *protocolo* (541 veces), *medida* (540 veces), *vida* (525 veces), *economía* (514 veces) e *reactivación* (486 veces). Asimismo, se pueden observar palabras alrededor como disciplina, trabajo, equipo, hassnassar, económico etc.

Figura 8. Nube de palabras de la muestra del 03 de julio de 2020.



Fuente: Twitter y elaboración propia en IRaMuTeQ y R.

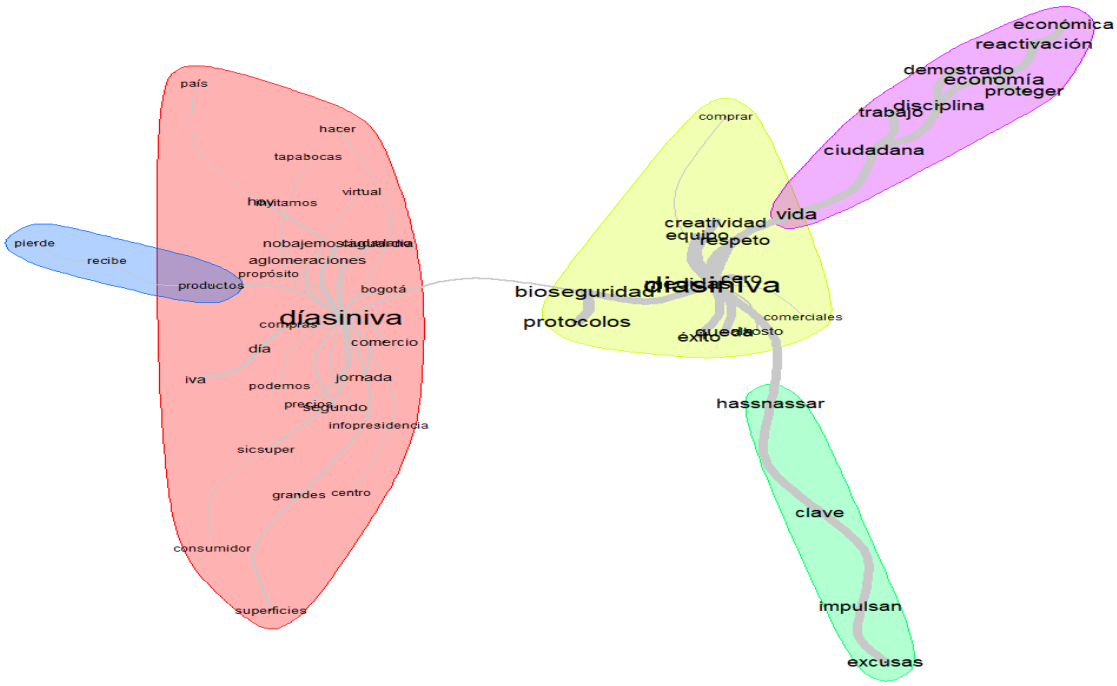
[8] Su única diferencia es la tilde en la palabra días.

En cuanto al análisis de similitud realizado (figura 9) con los textos analizados en el corpus, nos permite encontrar dos grandes grupos de palabras conectadas entre sí. *Diasiniva y Díasiniva.* Lo hemos representado con los halos para destacar los grupos de palabras conectadas y la relación que existe entre ellas y las diversas dependencias de las palabras que existen entre sí.

Ante las dos representaciones gráficas ya mencionadas, es de notar que en esta ocasión el discurso en torno al #diasiniva se vio mayormente marcado por el ámbito de la salud pública, sin dejar a un lado la reactivación económica como objetivo de la medida, pero enfatizando en gran medida en aquellas palabras de uso frecuente en medio de la emergencia sanitaria por COVID-19.

Cabe resaltar que ya para el 03 de julio de 2020, fecha en la que se genera esta segunda muestra y es realizado el segundo día sin IVA en el país, se tenían imágenes que construían un criterio frente a una experiencia previa con el primer día sin IVA, se tomaron acciones desde las administraciones públicas y los comercios para corregir algunas falencias en términos sanitarios observadas en la fecha anterior, más el primer pico de la pandemia amenazaba con el desborde de los servicios de salud y mostraba cifras alarmantes (hasta ese momento no vistas en el país) sobre contagios y muertes por COVID-19.

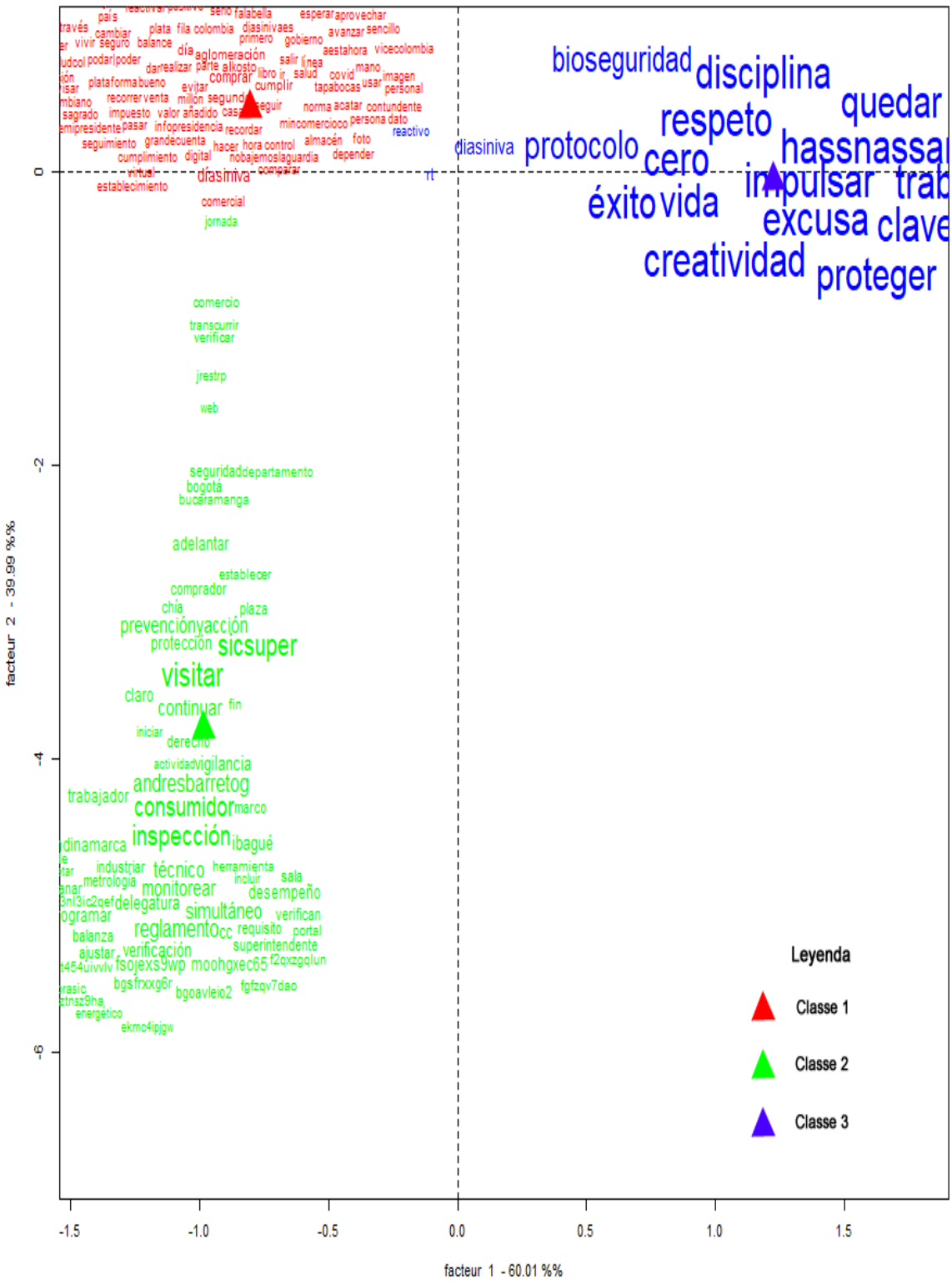
Figura 9. Análisis de similitud de los discursos analizados de la muestra del 07 de julio de 2020.



Revista Eletrônica Gestão e Serviços v.13, n. 1, pp. 3709 - 3730, Janeiro/Junho 2022.
ISSN Online: 2177-7284 e-mail: regs@metodista.br

Fuente: Twitter y elaboración propia en IRaMuTeQ y R.

Figura 10. Análisis de Especificidades y AFC de tres classes lexicales de la muestra del 03 de julio de 2020.



Fuente Twitter y elaboración propia en IRaMuTeQ y R.

En este sentido, se sigue con el análisis del AFC (figura 10) de la muestra del día 03/07/2020, la cual también resulto en tres classes: en el eje X están representados en el factor 1 por el 60,01 % y el Y con un factor de 39,99 % del cuerpo textual. La distribución de clases quedo de la siguiente la *classe* 1 quedo localizada en el eje positivos de X y en el negativo de Y de forma que se encuentra en el cuadrante superior derecho, la *classe* 2 quedo restringida a la parte negativa de X y del eje Y. Mientras que la *classe* 3 se encuentra en la parte positiva de X aunque distribuida de forma homogénea, tanto en la parte negativa, como la positiva del eje Y.

Como se construyó en la primera fecha, con el corpus textual de los tweets del 03/07/2020, realizamos la Clasificación Jerárquica Descendente (CDH) (figura 11) que formó un esquema jerárquico de *classes* de los vocabularios presentes en el corpus.

Figura 11. Dendograma de las tres classes lexicales obtenidas la clasificación jerárquica descendente de la muestra del 03 de julio de 2020.



Fuente Twitter y elaboración propia en IRaMuTeQ y R.

Mediante el CDH pudimos obtener el contenido de cada una de las classes de los textos analizados de los discursos, en este caso fueron 3 (tres) *classes*: la *classe* 1 con un 66,5 % hace referencia temas como diasiniva, comprar, alkosto, aglomeración, impuesto valor añadido, bioseguridad, etc. Dicha *classe* está relacionada con la *classe* 2 con un 5,7 % destacan las palabras, visitar, consumidor, reglamento, vigilancia, etc. La *classe* 3 con un 27,8% hace referencia a las palabras entre otras a; *impulsar, excusa, respeto, creatividad, trabajo, protección, etc.*

5 CONCLUSIONES

Luego de realizar un análisis pormenorizado de los datos recolectados de la red social de microblogueo Twitter con el #diasiniva para las dos fechas acontecidas durante el año 2020 (19 de junio y 03 de julio de 2020) bajo nubes de palabras, análisis de similitud, Clasificación Jerárquica Descendente y Análisis Factorial de Correspondencia, se encontró que los individuos utilizan las redes sociales, y en especial Twitter como escenario de encuentro ideal para expresión de ideas bajo el formato de mensajes cortos y en torno a la interacción de las representaciones individuales y colectivas respecto a un evento concreto.

Se observó cómo a pesar de tener una gran similitud entre las categorías y conjuntos de palabras de los corpus textuales recogidos en las muestras representativas para las fechas de los dos días sin IVA en Colombia, el análisis detallado bajo las diferentes herramientas de textometría genera un paisaje distinto de las representaciones de los discursos individuales y colectivos en cada una de ellas.

En congruencia con un contexto nacional cambiante en términos económicos, políticas y de salud pública entre las dos fechas, así mismo se observa una construcción y deconstrucción en las representaciones de las opiniones de los usuarios de Twitter frente al #diasiniva.

En este trabajo pudimos apreciar las posibilidades de trabajar con textometría para analizar tweets y establecer pautas de uso de los mismos frente a un evento concreto. Se pudo observar que es posible tratar datos de tweets con técnicas de textometría mediante el software IRaMuTeQ. En especial, el presente trabajo permite observar las posibilidades de esta herramienta como complemento a la evaluación de políticas económicas públicas. Dicho trabajo nos motiva a seguir esta línea de investigación para profundizar en esta metodología.

REFERÊNCIAS

ARRIETA BETTÍN, L; AVENDAÑO DE BARÓN, G. El discurso del tuit: un análisis lingüístico, sociodiscursivo y sociopragmático. **Cuadernos De Lingüística Hispánica**, (32), 107-130. 2018. https://doi.org/10.19053/0121053X.n32.2018.8123

BAVIERA, T. Técnicas para el Análisis de Sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. **Revista Dígitos**, 1(3), 33-50. 2017. Disponible en: https://revistadigitos.com/index.php/digitos/article/download/74/39. Acesso em: 11 jan. 2021.

BOUCHET-VALAT, M.; BASTIN, G. RcmdrPlugin. temis, a graphical integrated text mining solution in R. **The R Journal**, 5(1), p. 188-196. 2013

BUSÓN, C. La minería de opinión para el análisis del discurso de odio en las redes sociales / Opinion mining for the analysis of hate speech in social networks: Un estudio de caso sobre Paulo Freire en YouTube durante el periodo 2007-2019 / A case study on Paulo Freire on YouTube during the period 2007-2019. Commons. **Revista De Comunicación Y Ciudadanía Digital**, 9(1), 119-159. 2020. Disponible en: https://revistas.uca.es/index.php/cayp/article/view/5482. Acesso em: 10 maio. 2021.

CEPAL. Comisión Económica para América Latina y el Caribe. El COVID-19 y la crisis socioeconómica en América Latina y el Caribe. **Revista CEPAL.** n. 132, Edición Especial. p. 302. 2020

DIJK, T. A. VAN. El análisis crítico del discurso y el pensamiento social. Athenea digital, **Revista De Pensamiento E InvestigaciÓN Social**, 1(1), 18-24. 2002. doi:https://doi.org/10.5565/rev/athenead/v1n1.22

DIJK, T. A. VAN. El estudio del discurso. **En El discurso como estructura y proceso** Barcelona: Gedisa, p. 21-65. 2000.

GINI COLOMBIA, 2020. Disponível em: https://pt.countryeconomy.com/demografia/indice-de-gini/colombia

JUSTO, A. M; CAMARGO, B. V. **Estudos qualitativos e o uso de softwares para análises lexicais**. 2014

LAUER SA.*et al.* **The Incubation Period of Coronavirus Disease 2019 (COVID-19).** From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020; [Epub ahead of print 10 March 2020]. doi: https://doi.org/10.7326/M20-0504

LONGHI, J; SAIGH, D. **A textometrical analysis of french arts workers "fr. intermittents" on Twitter**. 2016. Disponible en: https://hal.archives-ouvertes.fr/hal-01359511/document.

MCALOON, C. G. **The incubation period of COVID-19** – A rapid systematic review and meta-analysis of observational research. 2020. medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20073957.

RATINAUD, P. **Visualisation chronologique des analyses ALCESTE**: application à Twitter avec l'exemple du hashtag# mariagepourtous. Actes des 12es Journées internationales d'Analyse statistique des Données Textuelles. Paris Sorbonne Nouvelle–Inalco. 2014. Disponible en;http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/46-JADT2014.pdf. Acesso em: 12 fev. 2021.

RATINAUD, P. **Étude d'un large corpus textuel avec IRaMuTeQ**: Twitter et le hashtag# mariagepourtous. Les représentations sociales. Théories, méthodes et applications, Louvain-la-Neuve, De Boeck Supérieur, p. 107-112, 2016.

REINERT, M. **Les "Mondes lexicaux" et leur "logique" à travers de l'analyse statistique d'un corpus de récits de cauchemars**. Langage et société, v. 66, n. 1, p. 5-39, 1993.

SARRICA, M., MINGO, I., MAZZARA, B., & LEONE, G. . The effects of lemmatization on textual analysis conducted with IRaMuTeQ: results in comparison. 13ème **Journées internationales d'Analyse statistique des Données Textuelles**. 2016. Disponible en: http://lexicometrica.univ-paris3.fr/jadt/jadt2016/01-ACTES/82897/82897.pdf. Acesso em: 07 mar. 2021.

SOUZA, MARLI; WALL, MARILENE; THULER, ANDREA; LOWEN, INGRID; PERES, AIDA. O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas. **Revista da Escola de Enfermagem da USP**. (faltam informações antes do número 52) 52. 2018. 10.1590/s1980-220x2017015003353.