

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ

ANÁLISIS ESTADISTICO DE REDES SOCIALES.

Taller 4: Analisis de Texto - JEP Caso 3

Cesar A Prieto Sarmiento - ceprieto@unal.edu.co

Alejandro Urrego Lopez - aurrego@unal.edu.co

23 de abril de 2024

1. INTRODUCCION Y CONTEXTO DEL CASO



La JEP o Jurisdicción Especial para la Paz es el componente de justicia del Sistema Integral de Verdad, Justicia, Reparación y no Repetición, creado por el Acuerdo de Paz entre el Gobierno Nacional y las FARC-EP. La JEP tiene la función de administrar justicia transicional y conocer los delitos cometidos en el marco del conflicto armado que se hubieran cometido antes del 1 de diciembre de 2016.

6.402 personas fueron presentadas en el país como guerrilleros dados de baja en combate cuando no lo eran. Esa fue la cifra preliminar que halló la Sala de Reconocimiento de Verdad entre 2002 y 2008, después de contrastar las versiones que han entregado los comparecientes sometidos ante la JEP con los expedientes de la Fiscalía, los informes de la Procuraduría, las cifras del Sistema Penal Acusatorio, el Observatorio de Memoria y Conflicto del Centro Nacional de Memoria Histórica y la Coordinación Colombia Europa Estados Unidos (CCEEU).

Esta aproximación de la JEP dentro del Caso 03, conocido como el de ‘falsos positivos’, arrojó que durante esos siete años se registró 78 % de toda la victimización y que los casos reportados pasaron de 792 en 2008 a 122 en 2009. O, por ejemplo, que 66 % de las víctimas se concentraron en 10 departamentos.

Antioquia, la Costa Caribe, Norte de Santander, Huila, Casanare y Meta son las seis zonas priorizadas en las que se concentran los esfuerzos de los magistrados que llevan el caso.

El Caso 3 de la JEP es titulado como **ASESINATOS Y DESAPARICIONES FORZADAS PRESENTADAS COMO BAJAS EN COMBATE POR AGENTES DEL ESTADO**. Se centra en 6 departamentos: **Meta** con 395 casos entre 2002 y 2008; **Casanare** con 248 casos entre 2002 y 2008; **Huila** con 274 casos entre 2002 y 2008; **Antioquia** con 1611 casos entre 2002 y 2008; **Norte de Santander** con 298 casos entre 2002 y 2008; y **Costa Caribe** con 538 casos entre 2002 y 2008. Sin embargo, para profundizar más, el perfil del caso 3 fue extraído directamente de la página web de la JEP también se encontró cual fue el criterio de priorización de las 6 zonas:

Las seis zonas, que entraron en la primera fase de la investigación del Caso 03, cuentan con elementos particulares de análisis. Es decir, fueron escogidas a partir de la contrastación de la información suministrada por entidades gubernamentales y la sociedad civil, además, las versiones que han entregado los comparecientes. No existe entre los seis subcasos un orden de prioridad. Además, el desarrollo, en cada una de las etapas, corresponde al avance propio de la investigación que depende de la disponibilidad de información y recursos.

Luego, para este documento se busca la aplicación de la metodología aprendida en clase con el propósito de llevar a cabo un análisis exhaustivo del discurso en el caso 03 de la JEP. Este análisis comprende la utilización de las técnicas como el análisis de sentimientos, así como el análisis de Bi-gramas y Skip-gramas. Se espera que a través de este enfoque metodológico, se puedan generar conclusiones de naturaleza estadística más amplias, enfocadas en el análisis del texto y las redes de palabras presentes en dicho discurso.

1.1. OBTENCION DEL TEXTO

Para la obtención de texto, se utilizaron principalmente dos bibliotecas de Python. La primera se llama `youtube transcript api`, y la segunda es `pytube`. Estas bibliotecas se implementaron en el código con la intención de obtener la transcripción automática de los videos de YouTube para los cuales existieran subtítulos. Por otra parte, para los videos que no tenían disponible los subtítulos, se implementó la alternativa propuesta por el profesor, encontrada en el repositorio de GitHub por el usuario `Dicklesworthstone` en el repositorio `bulk transcribe youtube videos from playlist` con esto, se pudo obtener la transcripción del texto de los 427 videos mencionados para el Caso 03 de la JEP.

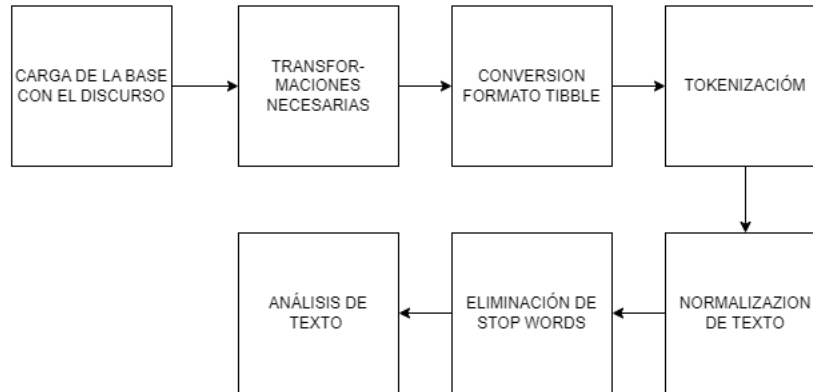
Por último, se realizó la concatenación de las transcripciones de los videos mencionados de forma ordenada para mantener un orden cronológico en la lista de reproducción del caso 03.

1.2. CARGA Y PREPARACIÓN DE LOS DATOS

La base resultante al obtener las transcripciones y unirlos como se indicó en la sección anterior, no está lista por sí sola para un análisis de texto, por lo cual se tuvieron que realizar varias transformaciones para de esta manera llegar a tener un objeto con el cual sea posible realizar el análisis textual que nos estamos proponiendo, dado esto y para entenderlo de una forma más sencilla se realizó el siguiente flujograma con la intención de dar a conocer el proceso por el cual fue expuesto el discurso del caso 03 para este análisis.

- Carga de la base
- Conversión en formato lista
- Transformación en formato tibble
- Tokenización
- Normalización de texto (Eliminación de números)
- Eliminación de Stop words

FLUJOGRAMA DEL PROCESADO DE LA BASE



Con todos estos cambios, se obtuvo al final una base con un total de 4.606.653 Tokens, los cuales están limpios de números, de stop words, y filtrados por las palabras que aparecen en un diccionario en español.

1.3. PALABRAS MÁS IMPORTANTES

Teniendo en cuenta los resultados obtenidos en el desarrollo anterior, lo primero que se propone es explorar la frecuencia con la que aparecen las palabras. Se obtuvo el total **n** señalando la cantidad de veces que la palabra fue encontrada en dentro de todo el discurso y por otra parte, también se obtuvo la frecuencia relativa **FreqC3** con la que estas palabras aparecen, con la intención para explorar la dimensionalidad más allá de solo tener el número de veces que aparece la palabra en el discurso.

	word	n
1	señor	64742
2	batallón	48272
3	comandante	37034
4	hechos	26184
5	doctor	24964
6	magistrado	24910
7	información	24268
8	operaciones	22965
9	brigada	21726
10	personas	21469

Tabla 1: Palabras más Frecuentes

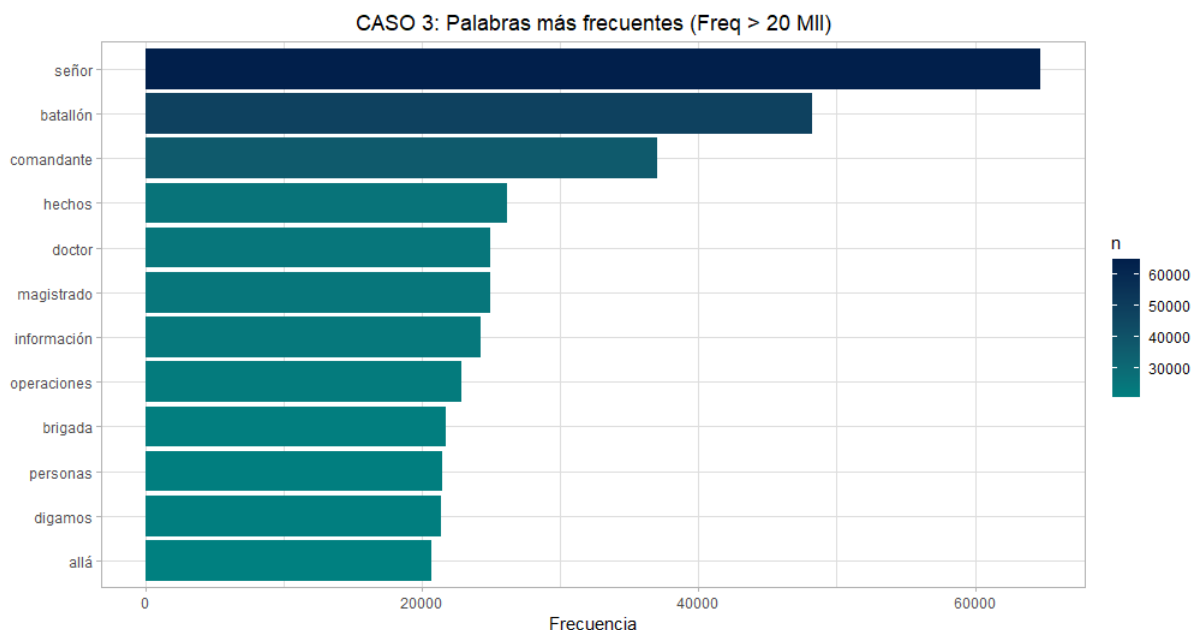
1

	word	Freq C3
1	señor	0.01447
2	batallón	0.01079
3	comandante	0.00828
4	hechos	0.00585
5	doctor	0.00558
6	magistrado	0.00557
7	información	0.00542
8	operaciones	0.00513
9	brigada	0.00486
10	personas	0.00480

Tabla 2: Frecuencia de la Palabras

2

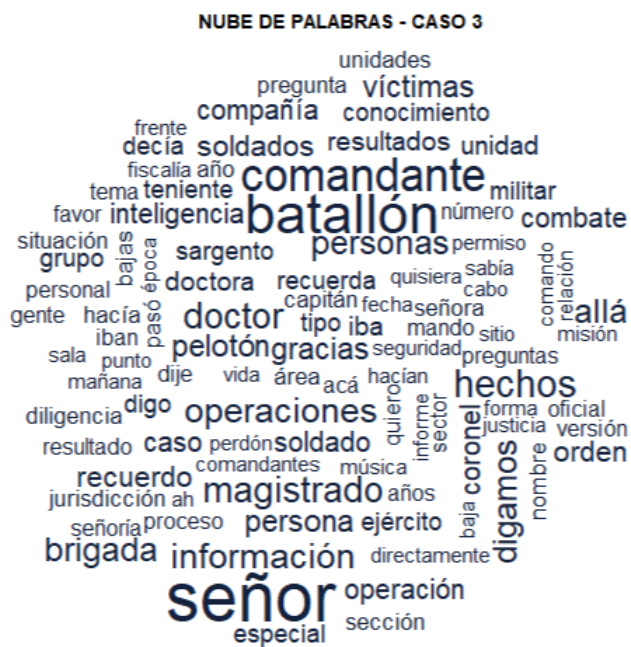
El histograma de frecuencias, además, sería visualizado de la siguiente manera al ser filtradas las palabras cuya frecuencia supere los 20 mil, revelando así una representación gráfica más precisa y detallada de la distribución de la frecuencia de palabras en el conjunto de datos analizado.



En las tablas 1 y 2 se muestran claramente las palabras más significativas en el análisis del texto, siendo la más frecuente la palabra *Señor* lo cual tiene sentido por el contexto militar y la forma respetuosa de dirigirse a otro o a un superior. Luego también se observan otras palabras que pertenecen a un contexto militar o político. Estas palabras pueden resultar siendo mucho más importantes más adelante cuando se realice un análisis más profundo utilizando agrupamiento u obteniendo otro tipo de medidas para clasificar la importancia de las mismas.

1.4. NUBE DE PALABRAS

Por otra parte, al realizar la nube de palabras se obtiene lo siguiente.

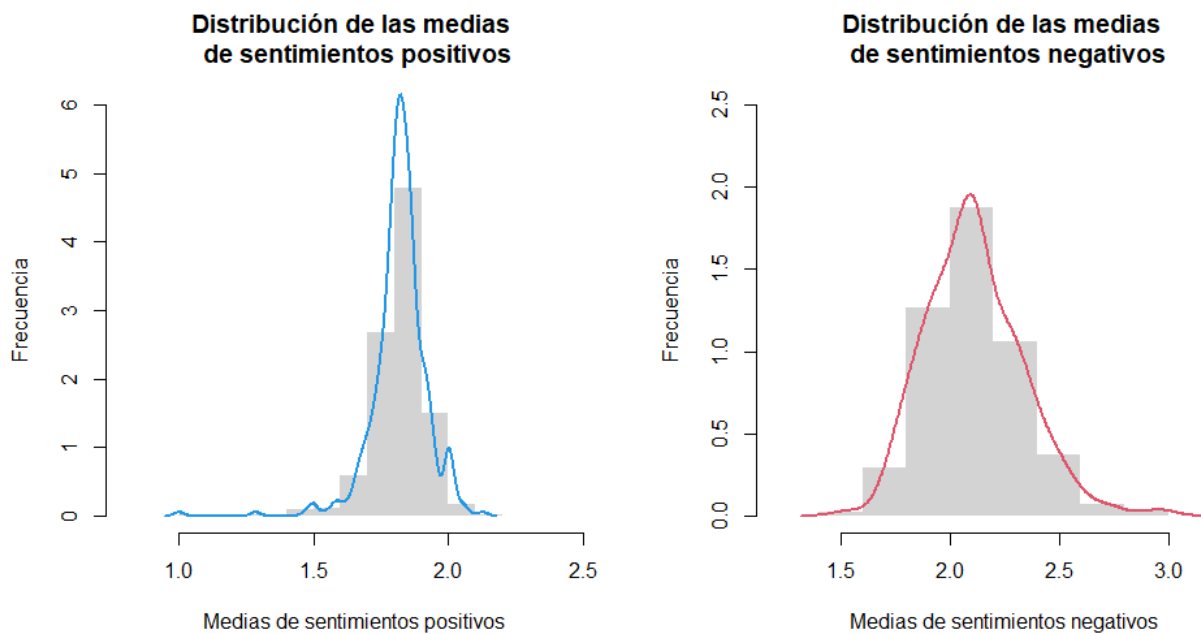


Al examinar la nube de palabras, también se observa un panorama notablemente político y militar, con ciertos temas destacándose prominentemente. Posteriormente, se evidenciará que estos temas desempeñan un papel significativo en la conversación del Caso 3, subrayando así su relevancia y trascendencia en el análisis.

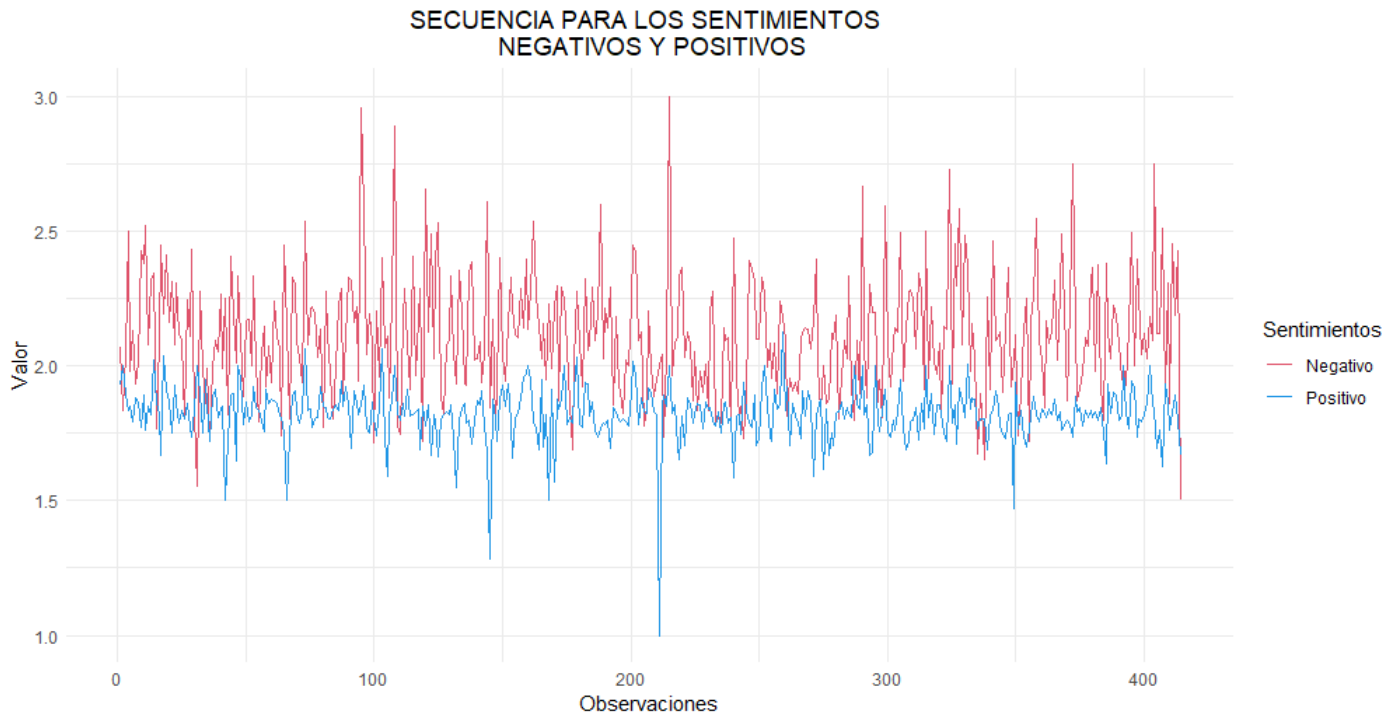
2. ANALISIS SENTIMIENTOS

Para el análisis secuencial dinámico de sentimientos, se deseaba realizar un análisis más exhaustivo, priorizando cada archivo de texto obtenido de cada uno de los videos en la lista de reproducción del caso 3. Con este propósito, se buscaba explorar de dos maneras distintas los sentimientos presentes en cada uno de los archivos. Antes que nada, se debe dejar en claro que para este desarrollo se utilizó la traducción del diccionario AFINN en inglés encontrado en el repositorio de Git (TAL). En primer lugar, se llevó a cabo un análisis a través de las medias de las puntuaciones encontradas en el diccionario AFINN, tanto para los sentimientos positivos encontrados y los negativos. En segundo lugar, se exploraron las puntuaciones mismas.

Para adentrarse más en los detalles de los métodos empleados en el análisis secuencial de sentimientos, se procedió a describir la primera forma que se utilizó. Esta consistió en emplear la implementación de los puntajes encontrados para el diccionario AFINN en las palabras del texto. Estos puntajes varían en un rango de -5 a 5, otorgando un valor negativo o positivo según la palabra exprese el sentimiento correspondiente, y siendo más cercano a 0 cuando la expresión del sentimiento es menos intensa.



La figura anterior muestra los histogramas de frecuencias relativas para las medias de los sentimientos positivos y negativos tomando ambas puntuaciones de 0 a 5, también muestra una línea superpuesta a cada histograma que representa la correspondiente estimación kernel de la densidad de las medias de los sentimientos. Por consiguiente, se puede apreciar como para los sentimientos positivos hay una acumulación más evidente en un punto mientras que para los sentimientos negativos esta acumulación no es tan grande, pero las medias parecen ser mayores que las de los sentimientos positivos.

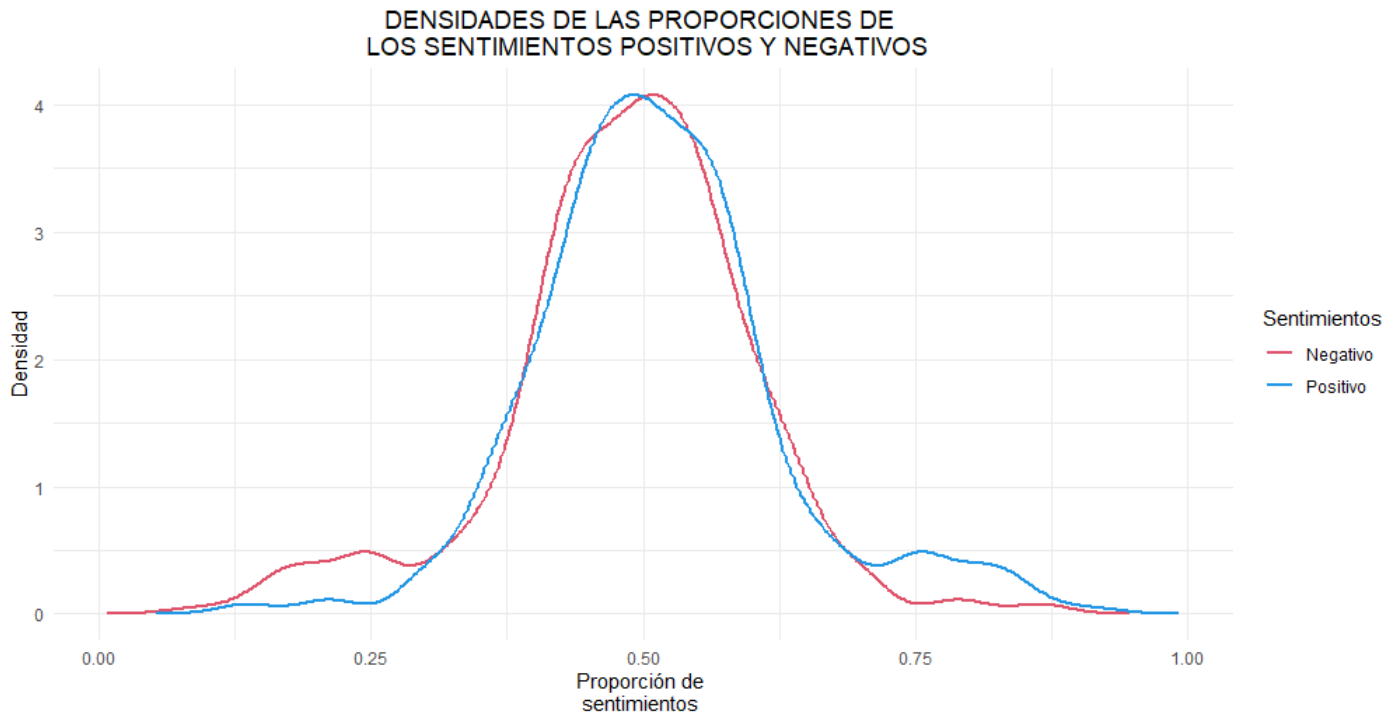


En la figura anterior se puede apreciar la secuencia de los sentimientos durante el discurso del caso 3 de la JEP, y con este gráfico se puede evidenciar como los sentimientos negativos son mayores a los positivos durante todo el discurso, en concordancia con lo dicho para el gráfico anterior los sentimientos positivos se ven más estables durante el discurso mientras que los negativos suelen tener cambios más bruscos.

Luego, se realizó una prueba de normalidad de **Shapiro-Wilk** para el vector de medias de los sentimientos positivos. El estadístico obtenido fue $W = 0.88555$, y el valor P resultó ser menor que $2.2e^{-16}$, lo que indica que el vector de medias de sentimientos positivos no es seguida por una distribución normal. Para el vector de medias de sentimientos negativos, también se llevó a cabo la prueba de normalidad de **Shapiro-Wilk**. El estadístico obtenido fue $W = 0.9818$, y el valor P fue igual a $4.529e^{-5}$, lo que sugiere que el vector de medias de sentimientos negativos no es seguida por una distribución normal. Lo anterior se realizó con el fin de conocer si metodológicamente sería correcto utilizar pruebas de hipótesis paramétricas que realicen comparaciones entre medias, como los resultados para ambos descartan una distribución normal se debe proseguir con pruebas no paramétricas para la comparación de medias.

A continuación, se llevó a cabo una prueba de **Mann-Whitney-Wilcoxon**, una prueba no paramétrica en este caso, dado que se descartó la normalidad de los datos. Se realizó un **Test de Wilcoxon** para comparar los vectores de las medias de los sentimientos negativos y positivos, con la hipótesis nula de que no hay diferencia entre los dos vectores de medias e hipótesis alternativa de que si hay diferencias entre ellos y las medias de los sentimientos negativos son mayores. El estadístico obtenido fue $W = 153526$, y el valor P es menor a $2.2e^{-16}$. Esto sugiere que hay suficiente evidencia para rechazar la hipótesis nula, lo que significa que se encontró diferencia significativa entre las distribuciones de los vectores de medias. Por lo tanto, se puede concluir que hay un cambio en la ubicación de las medias siendo la del grupo de sentimientos negativos mayor a la del grupo de sentimientos positivos.

Por otra parte, la segunda manera de explorar estos sentimientos en el discurso fue llevada a cabo mediante la obtención de las proporciones de sentimientos en cada uno de los archivos. Se hace la aclaración y se tiene en cuenta que las proporciones son simétricas, es decir, la proporción de positivos es la cantidad que le falta a la proporción de negativos para que en total sumen 1. Tras hacer esta aclaración, se obtuvo la proporción de sentimientos negativos y la proporción de sentimientos positivos para cada archivo, y a continuación se presentan las respectivas estimaciones de densidad kernel superpuestas, para observar como es el comportamiento de estas proporciones.



A partir de la gráfica anterior, se puede observar como las proporciones de sentimientos en cada parte del discurso del caso 3, se comportan de una forma muy similar siendo a veces superior la proporción de sentimientos negativos y a veces la proporción de sentimientos positivos. Lo cual no aporta significativamente al resultado anterior utilizando las medias. Por consiguiente, para este caso también nos proponemos a realizar pruebas de hipótesis para tener resultados más concluyentes.

Luego, muy similarmente al caso anterior se realizó la prueba de normalidad de **Shapiro-Wilk** para el vector de proporciones de los sentimientos positivos y también de los sentimientos negativos. Los estadísticos obtenidos son $W = 0.96519$ y $W = 0.96519$ los cuales por simetría de la proporción resultan ser iguales, por otro lado para el valor P resultó ser igual en ambos casos también con $P - value = 2.387e^{-8}$, lo que indica que tanto para el vector de proporciones de sentimientos positivos como para el de sentimientos negativos no son seguidos por una distribución normal. Como conclusión, se debe proseguir con pruebas no paramétricas para la comparación de proporciones

Considerando ahora lo anterior, se propuso al igual que con el ejercicio de las medias, llevar a cabo una la prueba de **Mann-Whitney-Wilcoxon**. Se realizó un **Test de Wilcoxon** para comparar los vectores de las proporciones de los sentimientos negativos y positivos, con la hipótesis nula de que no hay diferencia entre los dos vectores de medias e hipótesis alternativa de que si hay diferencias entre ellos y las medias de los sentimientos negativos son mayores. El estadístico obtenido fue $W = 93901$, y el valor P es igual a 0.008569 . Esto sugiere que hay suficiente evidencia para rechazar la hipótesis nula, lo que significa que se encontraron diferencia significativa entre las distribuciones de los vectores de proporciones a favor de las proporciones de los sentimientos negativos. Por lo tanto, se puede concluir que, la proporción de sentimientos negativos a través del discurso del caso 3 es significativamente superior a la de los sentimientos positivos.

3. ANÁLISIS DE BIGRAMAS

Para generar los bigramas, se toman todos los tokens y se añade el token siguiente. Para una mayor cohesión en este caso, se eliminan los acentos de los bigramas y se calcula su frecuencia.

3.1. DEFINICIÓN DE LA RED A PARTIR DE LOS PESOS

Ahora se necesita definir un umbral para crear la red asociada con los bigramas generados a partir del caso 03. Para definir este umbral, se utilizará el cuantil de las frecuencias de los bigramas, ya que se desea un número específico de palabras para evitar exceder la capacidad computacional disponible. De esta manera, podremos observar cuántas veces se repiten los términos seleccionados.

Si se desea, se puede reducir el umbral para obtener más bigramas para analizar en la red. Sin embargo, en este caso se busca tener aproximadamente 200 nodos en la red. Por lo tanto, para encontrar el cuantil que nos dé al menos 200 bigramas, podemos expresarlo como:

$$\alpha = 1 - \frac{200}{n}$$

Donde α representa el cuantil y n es el número total de bigramas.

Esto se realiza debido a dos razones. La primera es que, debido a la cantidad de videos, la densidad de la red resultante al definir un umbral pequeño sería muy grande y poco informativa. Además, con 200 nodos, los algoritmos de agrupamiento no requieren tanto poder computacional. Así pues, se trabaja con un umbral de $n = 400$.

Después de obtener los bigramas, se genera una red donde las palabras son los nodos y los bigramas son las aristas. A esta red se le calculan varias medidas descriptivas, se grafica y luego se analiza su componente conexas.

Medida	Valor
Distancia media	4.885
Grado medio	1.854
Desviación del grado	2.177
Número de clanes	3
Densidad	0.008
Transitividad	0.040
Asortatividad	0.064

Figura 1: Medidas descriptivas de la red generada con bigramas.

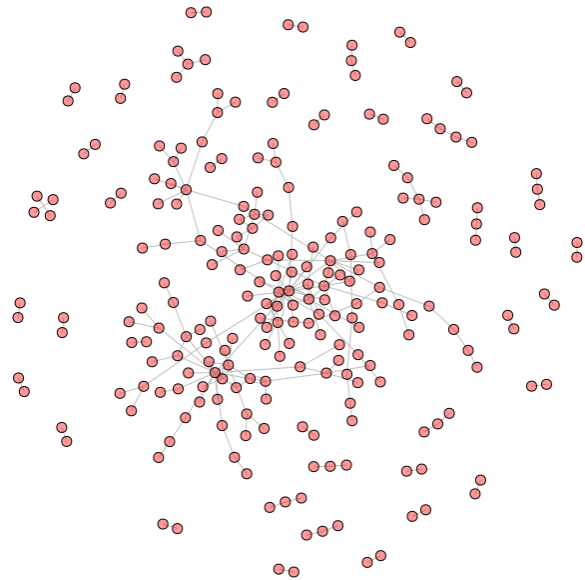


Figura 2: Bigrama

Se puede observar que la distancia geodésica promedio es alrededor de 4. Esto quiere decir que, en promedio, se necesitan 4 aristas para llegar de una palabra a otra por medio de las aristas.

El grado medio es de 1.85. Esto quiere decir que el número de aristas incidentes por palabra en promedio es de 1.85, es decir, que en promedio, las palabras de la red están conectadas con 1.85 palabras.

La desviación estándar del grado es de 2.18. Al ser mayor que la media, refleja que los datos son muy variables y que el grado cambia mucho entre ellos.

El número de clan es de 3. Esto quiere decir que el subgrafo que son completos (todas las palabras están conectadas entre sí) es 3, lo que indica una conectividad baja de la red.

La transitividad global del grafo, que se calcula como el número de tríadas cerradas sobre el número de tríadas (siendo una tríada cerrada cualquier subgrafo de 3 palabras completo), ayuda a ver que, por lo general, las palabras adyacentes entre sí no suelen ser adyacentes a otra palabra en común.

La asortatividad es un indicador de homofilia, es decir, qué tan parecidos son los grados de las palabras que están conectadas. Se puede observar, por su bajo valor, que no son muy parecidos.

3.2. COMPONENTE CONEXA

Con la componente gigante del grafo, se procede a hacer clustering. Se analiza cuál de los métodos de agrupamiento tiene una mayor modularidad, la cual mide qué tan buena es la división o qué tan separados están los diferentes tipos de vértices entre sí. Por lo tanto, se escoge la partición que surge del método Edge-betweenness.

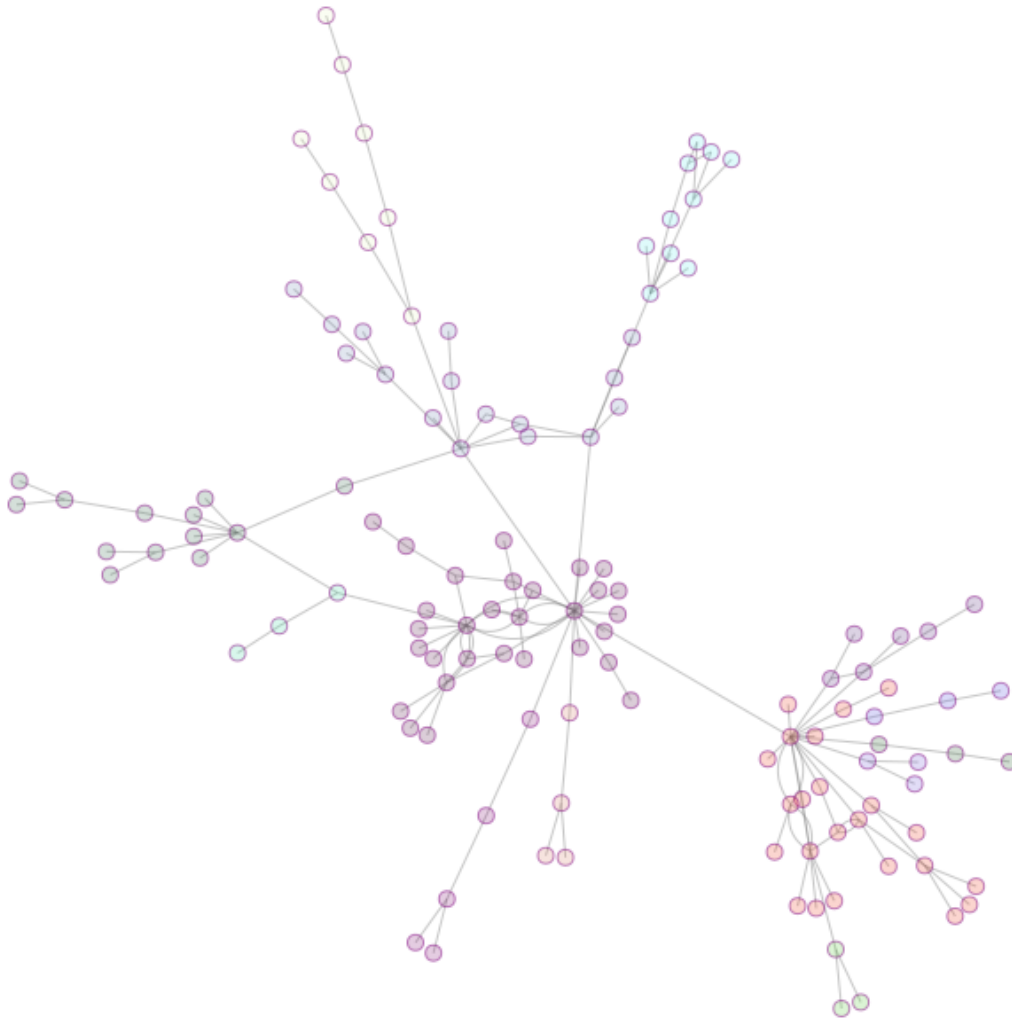


Figura 3: Cluster Componente Gigante Bigrama

Se puede observar que hay 14 clústeres. En cada clúster, se organizan las palabras en orden descendente teniendo en cuenta sus medidas de centralidad. Posteriormente, se utiliza GPT-4 para determinar cuál es el tema que más relaciona estas palabras, lo que permite identificar los temas más importantes que se trataron en el caso 03.

Cluster	Palabras
0	combate muertes bajas muertos baja muerte presentadas legitimamente ilegitimamente ilegitimas
1	operaciones inteligencia seccion oficial jefe orden desarrollo ordenes informacion anexo publico batalla dia operacion pago ministerio
2	jurisdiccion especial paz grupo
3	batallon comandante brigada comando comandantes trasladado contraguerrillas panza comante contraguerrilla pigoanza magdalena guerrilla compania ejercito peloton directamente abrigada ejecutivo division novena cuarta movil operativo sali nacional c b instruccion nivel
4	coronel teniente torres pinto suarez escalante
5	señor gracias recuerdo dias magistrado preguntas hechos claro procurador pregunta compareciente doctor nombre muchisimas doctora dios exactamente honorable permiso sitio recuerda completo acuerdo
6	militares fuerzas unidades especiales tacticas grupos misiones armados
7	infanteria numero cedula identificado ciudadania
8	area militar penal vida control servicio escuela carrera justicia hoja folio restaurativa ordinaria
9	voluntaria version diligencia
10	jose san agustin
11	señora magistrada procuradora
12	general montoya nacion
13	unidad tactica mision

Tabla 3: Clusters y palabras asociadas

Así se puede decir que el tema asociado a cada cluster es:

Número de Cluster	Tema o Tópico Principal
0	Conflictos y bajas militares
1	Operaciones militares y estructura de mando
2	Jurisdicción especial para la paz
3	Estructura y organización militar
4	Oficiales militares
5	Procedimientos judiciales y testimonios
6	Fuerzas especiales y tácticas militares
7	Identificación militar y registro personal
8	Sistema penal y justicia militar
9	Procesos voluntarios y diligencias legales
10	Referencias personales o geográficas
11	Figuras legales y judiciales
12	Alto mando militar
13	Tácticas y misiones militares

Tabla 4: Clusters y sus Temas o Tópicos Principales

4. ANÁLISIS DE SKIP-GRAMAS

Para generar los skipgramas, se toman todos los tokens y se añade el token siguiente al token que precede. Posteriormente, se unen con un espacio. Luego, todos los procedimientos son los mismos que los realizados con los bigramas.

4.1. DEFINICIÓN DE LA RED A PARTIR DE LOS PESOS

Para el umbral trabajado con los skip-gramas se tomó un umbral de $n=400$, se grafica la red y se calculan algunas medidas descriptivas de la misma:

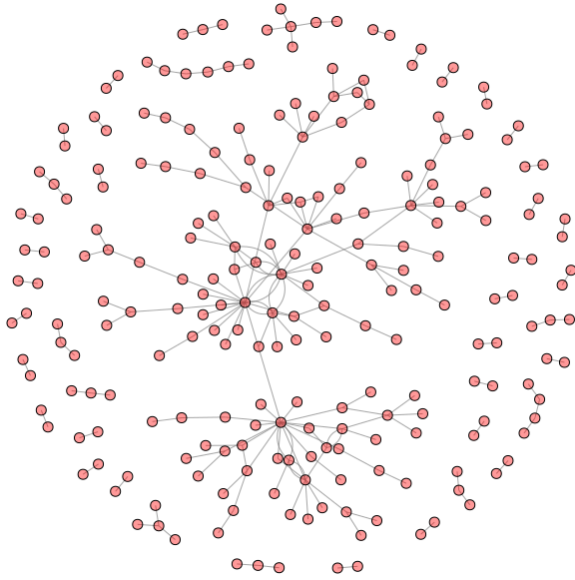


Figura 4: Skip-grama

Medida	Valor
Distancia media	4.885
Grado medio	1.854
Desviación del grado	2.178
Número de clan	3
Densidad	0.008
Transitividad	0.040
Asortatividad	0.064

Figura 5: Medidas del grafo

Al observar las medidas descriptivas, se puede notar que son muy parecidas a las estadísticas generadas por los bigramas; por lo tanto, se llega a las mismas conclusiones.

4.2. COMPONENTE CONEXA

4.2.1. ANÁLISIS DE COMPONENTE CONEXA

Con la componente gigante del grafo, se procede a hacer clustering. Se analiza cuál de los métodos de agrupamiento tiene una mayor modularidad, Por lo tanto, se escoge la partición que surge del método Edge-betweenness.

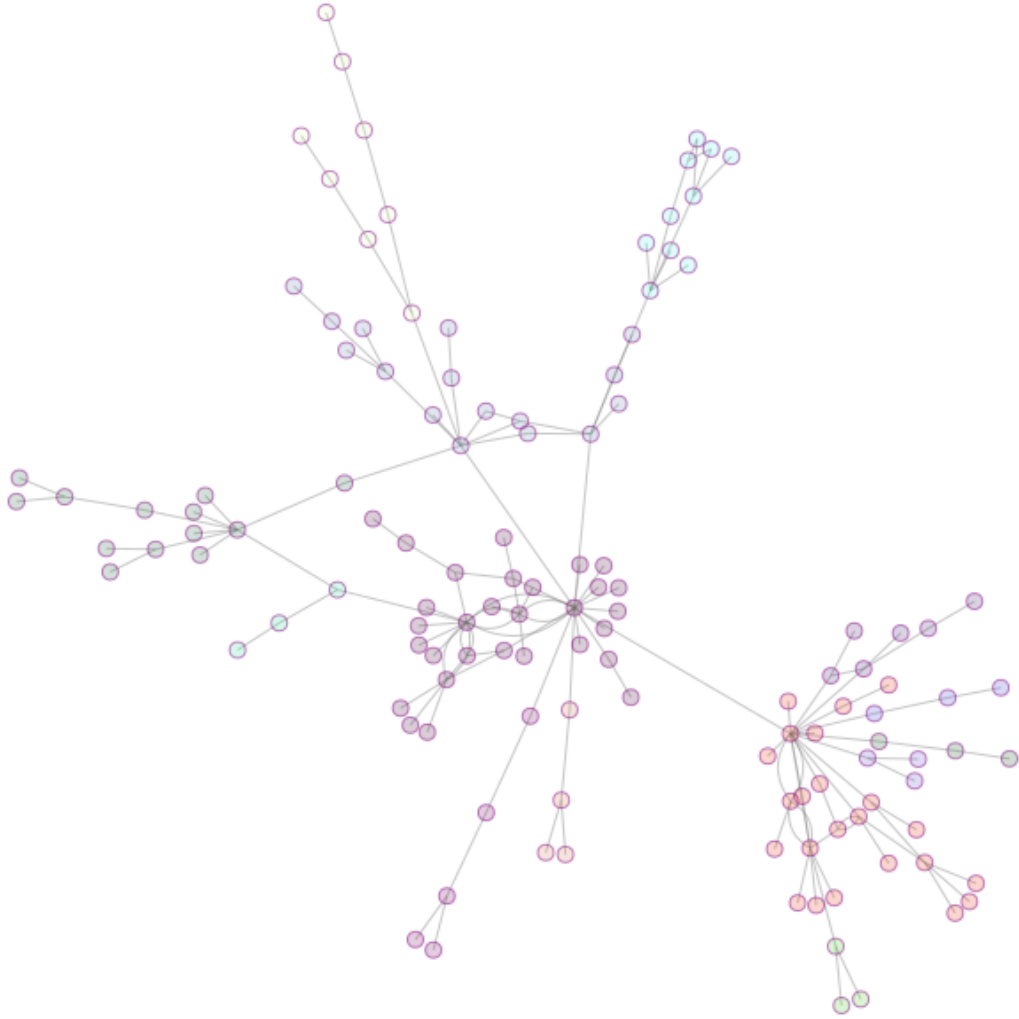


Figura 6: Cluster Componente Gigante Bigrama

Se puede observar que hay 14 clústeres. Siguiendo el mismo procedimiento para el Bigrama se obtiene lo siguiente:

Cluster	Tema o Tópico Principal
0	Conflictos y bajas militares
1	Operaciones militares y de inteligencia
2	Jurisdicción especial para la paz
3	Estructura y movimientos militares
4	Rangos y personal militar
5	Procedimientos legales y judiciales
6	Fuerzas y unidades militares
7	Identificación militar
8	Instituciones y vida militar
9	Procesos voluntarios y diligencias
10	Nombres propios
11	Autoridades judiciales
12	Altos mandos y liderazgo militar
13	Estrategias y misiones tácticas

Tabla 5: Clusters y sus Temas o Tópicos Principales

Cluster	Palabras
0	combate muertes bajas baja muerte muertos presentadas legitimamente ilegitimamente ilegitimas
1	operaciones inteligencia seccion oficial jefe orden desarrollo ordenes informacion anexo publico dia batalla operacion pago ministerio
2	jurisdiccion especial grupo paz
3	batallon comandante brigada comando comandantes trasladado guerrilla panza contraguerrilla magdalena contraguerrillas comandante pigoanza compañía ejercito peloton division directamente abrigada ejecutivo movil cuarta novena operativo sali nacional b instruccion c nivel
4	coronel teniente torres pinto suarez escalante
5	señor gracias recuerdo dias magistrado preguntas hechos claro pregunta procurador compareciente doctor nombre dios muchisimas doctora exactamente honorable permiso sitio recuerda completo acuerdo
6	militares fuerzas unidades especiales tacticas grupos misiones armados
7	infanteria numero cedula identificado ciudadania
8	area militar penal vida control carrera escuela servicio justicia hoja folio restaurativa ordinaria
9	voluntaria version diligencia
10	jose san agustin
11	señora magistrada procuradora
12	general nacion montoya
13	unidad tactica mision

Tabla 6: Clusters y palabras asociadas

Se puede observar que las palabras de cada clúster son similares tanto en las componentes gigantes de los bigramas como en los skip-gramas. Por lo tanto, los temas o tópicos principales de los vídeos son similares, lo que brinda aún más respaldo a la relevancia de estos. Es algo bastante útil y a tener en cuenta, debido a la relevancia social y política que tiene este tema en Colombia.

5. CONCLUSIONES

En conclusión, los resultados del análisis de texto realizado para el discurso del caso 03 revelan varias lecciones importantes. En primer lugar, se observa cómo predominan los sentimientos negativos sobre los positivos a lo largo del discurso. Esta tendencia cobra sentido dada la naturaleza del caso 03 en el país, el cual se centra en las desapariciones forzadas perpetradas por las autoridades, también conocidas como falsos positivos. En este contexto, es comprensible que un evento tan trágico, que ha dejado una profunda huella de muerte en la nación, se refleje en un discurso donde predominen los sentimientos negativos.

La tristeza, el dolor, la angustia, la ira, el enojo y la incertidumbre son algunos de los sentimientos que se destacan a lo largo del discurso, opacando en gran medida cualquier momento de positividad. Sin embargo, es importante destacar que, aunque estos sentimientos negativos son dominantes, no se niega la existencia de emociones positivas que también desempeñan un papel significativo en el contexto del caso 03. Entre estos se encuentran la esperanza y el perdón, expresados por parte de los afectados, lo que añade complejidad al análisis y refleja la diversidad de respuestas emocionales frente a situaciones de crisis y trauma.

En cuanto a los análisis realizados sobre la componente conexa o la componente más grande para los Bi-Gramas y los Skip-Gramas, se han obtenido resultados notablemente similares. Entre los hallazgos más relevantes que desean resaltar se encuentran los clústeres o grupos identificados en las redes de palabras. Como se pudo observar en las tablas 3 y 6, el algoritmo utilizado para agrupar las palabras tuvo en cuenta sus medidas de centralidad, logrando capturar esencialmente grupos de palabras relacionadas por temáticas comunes.

Estas temáticas, comprendidas por conflictos, situaciones militares y/o políticas, tampoco sorprenden al considerar el contexto del caso 03 y la trayectoria del discurso. Es evidente que estas temáticas están intrínsecamente relacionadas con el contenido del discurso y el contexto circundante.

Por lo tanto, se puede concluir que el discurso del caso 03 de la JEP tiene una connotación sentimental negativa, con tematica militar, política y social, lo cual nos invita a reflexionar sobre el contexto nacional y a considerar cómo podemos prevenir la repetición de eventos similares en el futuro. Este análisis resalta la importancia de comprender y abordar adecuadamente los problemas subyacentes a nivel militar, político y social para promover una sociedad más justa y pacífica.

6. BIBLIOGRAFIA

- Jurisdicción Especial para la Paz (JEP). Sitio web oficial. Recuperado de: <https://www.jep.gov.co/JEP/Paginas/Jurisdiccion-Especial-para-la-Paz.aspx>
- JEP - Caso 03. Sitio web oficial. Recuperado de: <https://www.jep.gov.co/macrocasos/caso03.html#container>
- Lista de Reproducción del Caso 03 de la JEP. [Vídeo en línea]. Recuperado de: https://www.youtube.com/playlist?list=PLbtegW3d3L4Id_lrAgBRDGf-k_bwEYH6_
- Repositorio GitHub de *bulk_transcribe_youtube_videos_from_playlist*. Recuperado de: https://github.com/Dicklesworthstone/bulk_transcribe_youtube_videos_from_playlist
- AFINN en español. Repositorio GitHub. Recuperado de: <https://github.com/jboscomendoza/lexicos-nrc-afinn>
- Sosa, J. (2022). Cuaderno Rpubs: Análisis de datos sobre [tema del cuaderno]. Rpubs. Recuperado de: <https://rpubs.com/jstats1702/946584>