



Metodologías de análisis de entrevistas utilizando Procesamiento de Lenguaje Natural (PLN)

Grupo de procesamiento de lenguaje Natural

Comisión de la Verdad
SIM - Analítica - Procesamiento de Lenguaje Natural
Bogota, Colombia
2022

Contenido

1 Métricas Estructurales	4
1.1 Análisis de redes de entrevistas de masacres	5
1.2 Análisis de redes de relaciones presentes en el corpus de entrevistas de Justicia y Paz	11
1.3 Análisis de redes de entrevistas de dinámicas del conflicto	18
1.4 Análisis de redes de entidades presentes en entrevistas etiquetadas por el equipo de resistencia	25
1.5 Análisis de redes de entidades presentes en sentencias de la Unidad de Restitución de Tierras (URT) Nariño	31
1.6 Análisis de redes de entidades presentes en sentencias de la Unidad de Restitución de Tierras (URT) Cesar	38
Bibliografía	43
2 Relaciones y Coocurrencias	45
2.1 Análisis de entramado sobre las entrevistas colectivas	46
2.2 Análisis de entramado sobre las entrevistas a sujetos colectivos	52
Bibliografía	57
3 Dimensiones	60
3.1 Análisis automatizado de corpus: indicialidad y dimensión de sevicia utilizando herramientas de procesamiento del lenguaje natural	61
Bibliografía	67
4 Clasificadores, intensidades y lexicones	69
4.1 Ejecuciones extrajudiciales en el conflicto armado colombiano	70
4.2 Duelo en los familiares de desaparecidos en Colombia	77
4.3 Análisis de estigmatización de comunidades étnicas en Buenaventura	85
Bibliografía	94
5 Lexicones	96
5.1 Análisis del lenguaje afectivo utilizado en entrevistas	96

Bibliografía	108
6 Frecuencias por temas y ministerios	110
6.1 Análisis de los discursos presidenciales de posesión en Colombia según términos referentes a temáticas y ministerios	111
Bibliografía	118
7 Sistema de recomendación	120
7.1 Análisis de sistema de recomendación de entrevistas etiquetadas por el equipo de resistencia	122
Bibliografía	127
8 Conclusión	129

Lista de Figuras

1-1	Organizaciones-Verbos	6
1-2	Organizaciones-Verbos	12
1-3	Organizaciones-Verbos	19
1-4	Chunks - Lugares	26
1-5	Organización - Organización	32
1-6	Organización - Organización	39
2-1	Frecuencia de sustantivos - adjetivos	48
2-2	Network lecture Organizaciones - Organizaciones	50
2-3	Diagrama Organizaciones - Organizaciones	51
2-4	Frecuencia de organizaciones	54
2-5	Network lecture personas - personas	55
2-6	Diagrama personas - personas	56
3-1	Histórico de tipos de sevcia	62
3-2	Análisis histórico dimensión de la sevcia (1939 - 2020)	64
3-3	Promedio de dimensiones respecto a la muestra	64
3-4	Frecuencia de los términos se sevcia en las entrevistas	65
3-5	Promedio índices sevcia por año	66
3-6	Caja y Bigotes: Fuerza pública	67
4-1	Verbos más frecuentes	72
4-2	Nube de palabras: verbos más frecuentes	73
4-3	FastText: Homicidios	74
4-4	Ejecuciones Extrajudiciales: Valores referentes a víctimas	75
4-5	Ejecuciones Extrajudiciales: Caja y Bigotes Víctima	76
4-6	Adjetivos más frecuentes	79
4-7	Nube de palabras de los adjetivos más frecuentes	80
4-8	Fasttext: Desapariciones Forzadas	82
4-9	Análisis histórico dimensión del duelo (1931 – 2019) Categoría de Búsqueda	83
4-10	Duelo en categoría de Búsqueda: Cajas y Bigotes	84
4-11	Gráfica de los 50 verbos más frecuentes	88
4-12	Nube de palabras de los 50 verbos más frecuentes	89
4-13	Gráfica modelo Fast Text	90

4-14	Gráfica de análisis histórico del promedio de frecuencia de términos relacionados con la estigmatización verbal	91
4-15	Gráfica de promedio de intensidad de los términos de la muestra	92
5-1	Categorías emocionales en “Tortura y otros tratos crueles”	98
5-2	Categorías emocionales en “Detención arbitraria”	100
5-3	Categorías emocionales en “Homicidio”	101
5-4	Categorías emocionales en “Desplazamiento forzado”	103
5-5	Categorías emocionales en “Ataque a bien protegido”	104
5-6	Categorías emocionales en “Desaparición forzada”	105
5-7	Intensity NCR en “Exilio-Homicidio”	106
5-8	Arousal Anew en “Exilio-Homicidio”	107
6-1	Álvaro Uribe (2002-2006) - Frecuencia de términos por ministerio en el discurso	114
6-2	Álvaro Uribe (2002-2006) - Frecuencia de términos por tema en el discurso .	115
6-3	Juan Manuel Santos (2010-2014) - Frecuencia de términos por ministerio en el discurso	116
6-4	Juan Manuel Santos (2010-2014) - Frecuencia de términos por tema en el discurso	117
7-1	Grafo de relaciones para Resistencia n=6, lift=1, confidence=0.1	123
7-2	Grafo de relaciones para Resistencia n=16, lift=0.6, confidence=0.1	126

Lista de Tablas

1-1	All Degree: Organizaciones - Verbos	7
1-2	Betweenness centrality: Organizaciones - Verbos	8
1-3	All closeness centrality: Organizaciones - Verbos	9
1-4	Hubs: Organizaciones - Verbos	9
1-5	Authorities: Organizaciones - Verbos	10
1-6	All Degree: Organización - Organización	14
1-7	Betweenness centrality: Organización - Organización	15
1-8	All closeness centrality: Organización - Organización	16
1-9	Hubs: Organización - Organización	16
1-10	Authorities: Organización - Organización	17
1-11	All Degree: Organización - Verbo	20
1-12	Betweenness centrality: Organización - Verbo	21
1-13	All closeness centrality: Organización - Verbo	22
1-14	Hubs: Organización - Verbo	23
1-15	Authorities: Organización - Verbo	23
1-16	All Degree: Chunks - Lugares	27
1-17	Betweenness Centrality: Chunks - Lugares	28
1-18	All Closeness Centrality: Chunks - Lugares	29
1-19	Hubs: Chunks - Lugares	29
1-20	Authorities: Chunks - Lugares	30
1-21	All Degree: Organización - Organización	33
1-22	Betweenness centrality: Organización - Organización	34
1-23	All closeness centrality: Organización - Organización	35
1-24	Hubs: Organización - Organización	35
1-25	Authorities: Organización - Organización	36
1-26	All Degree: Organizaciones - Organizaciones	40
1-27	Betweenness centrality: Organizaciones - Organizaciones	40
1-28	All closeness centrality: Organizaciones - Organizaciones	41
1-29	Hubs: Organizaciones - Organizaciones	41
1-30	Authorities: Organizaciones - Organizaciones	42
5-1	Categorías emocionales en “Tortura y otros tratos crueles”	98
5-2	Categorías emocionales en “Detención arbitraria”	100

5-3	Categorías emocionales en “Homicidio”	102
5-4	Categorías emocionales en “Desplazamiento forzado”	103
5-5	Categorías emocionales en “Ataque a bien protegido”	104
5-6	Categorías emocionales en “Desaparición forzada”	106
7-1	Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=1	123
7-2	Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=0.6	124
7-2	Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=0.6	125

Introducción

A lo largo de la creación de la Comisión de la Verdad, el grupo de Procesamiento de Lenguaje Natural, para el presente libro, se ha encargado del desarrollo de quince (15) diferentes informes redactados en congruencia con problemáticas acontecidas dentro del conflicto armado colombiano. Esto fue posible gracias a la recolección de información encontrada dentro de los diferentes corpus que contiene la Comisión con las etiquetas del metabuscador bajo múltiples categorías. Estas se componen de las diversas entrevistas a las poblaciones objetivo de cada informe.

Para el análisis de cada uno de estos informes, fue necesaria la aplicación de diferentes metodologías de acuerdo con los diversos intereses en la recolección, procesamiento y análisis de la información. No obstante, al encontrarse objetivos parecidos entre algunos informes, estos compartieron la metodología aplicada.

Gracias a lo explicado de manera breve anteriormente, el objetivo general de este libro es la presentación, análisis y explicación de cada una de las metodologías aplicadas en la temporalidad de realización de cada uno de los informes del grupo de analítica.

La división de este libro se desarrolla en siete capítulos junto con la conclusión general. Cada capítulo comprende una metodología diferente con la cantidad n de informes que utilizan dicha metodología. A continuación, se esbozará de manera breve la comprensión de cada capítulo por cada enumeración de acuerdo con la presentación de sus resultados.

1. Presentación de resultados mediante grafos desarrollados en VosViewer y tablas de métricas estructurales.
2. Presentación de resultados mediante grafos desarrollados en VosViewer, gráficas de frecuencia junto con su traducción en nubes de palabras y gráficos heatmap.
3. Presentación de resultados mediante gráficas de barras que ejemplifican frecuencias e índices gracias a valores de promedios anuales y gráficas de cajas de bigotes.
4. Presentación de resultados mediante gráficas de frecuencia, nubes de palabras, modelo de graficación FastText, gráficas de barras de centralización de datos y cajas de bigotes.

5. Presentación de resultados mediante gráficas de intensidad por pares, gráficas de frecuencia por barras y tablas de frecuencia.
6. Presentación de resultados mediante dos tipos de gráficas de frecuencia de acuerdo a categorías establecidas.
7. Presentación de resultados mediante tablas del sistema de recomendación y grafos de relaciones producto del sistema de recomendación.

1 Métricas Estructurales

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las sentencias de los diferentes corpus utilizados, dependiendo de los objetivos de los diversos informes, para la construcción de redes que apunten a relaciones relevantes y/o interesantes. Los resultados de las redes son evaluados a partir de cuatro distintas métricas estructurales:

- All degree: Señala los términos más y menos importantes en el corpus con su valor de frecuencia, es decir, representa un número de enlaces que se tiene con otros nodos [1]. Su medida se calcula con la siguiente expresión:

$$C_D(i) = \frac{1}{2} \sum_{i=1}^N k_i$$

Dónde k_i representa el nodo existente dentro de la red.

- Betweenness centrality: Ordena los nodos determinando la centralidad de todos los nodos dada una pareja ordenada de nodos [1], su medida se calcula con la siguiente expresión:

$$C_B(i) = \sum_{j < k} \frac{p_{jk}(i)}{p_j k}$$

Donde p_{jk} son los términos que representan la fracción de caminos cortos que atraviesan el nodo x , y $p_j k$ representa todos los caminos cortos entre cada nodo.

- All closeness centrality: Evalúa todas las distancias de longitud de cada nodo, que tan lejos o que tan cerca están unos de otros, cuáles términos pueden llegar a estar más lejos y resalta los nodos más centrales de toda la red [1], su medida se calcula con la siguiente expresión:

$$C_c(i) = \frac{n - 1}{\sum_{j=1}^n d_{i,j}}$$

Donde $n - 1$ Es la suma de todos los nodos existentes, y $d_{i,j}$ representa todos los caminos cortos entre cada nodo.

- Hubs and Authorities: Denota el valor de circulación de información en torno a una entidad y las personas mencionadas dentro de la red [1].

La presentación de estos análisis se desarrolló mediante network lectures hechos por medio de la herramienta VosViewer. Dentro de estos grafos se puede destacar la relevancia de los términos de las parejas de nodos examinadas de acuerdo con los diferentes análisis de los informes que comparten este tipo de metodología. Entre mayor sea el radio de las bolas que contienen los términos, mayor será su presencia. Del mismo modo, los resultados son presentados mediante tablas que permiten evidenciar los ítems de los diferentes corpus analizados usando las métricas estructurales mencionadas previamente.

A continuación, se presentarán secciones respecto a los diferentes informes que utilizan esta metodología dentro de sus análisis

1.1. Análisis de redes de entrevistas de masacres

Este informe fue culminado el 26 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Dentro de este, se detalla el trabajo de procesamiento de corpus de las entrevistas recolectadas por la Comisión de la Verdad y clasificadas bajo la etiqueta de Masacres. A lo largo de todo el informe se tiene en cuenta que las masacres hacen parte y ejemplifican una categoría dentro de un hecho victimizante del conflicto armado y su relevancia en el esclarecimiento de la verdad y la no repetición de este. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Chunks - Adjetivos
- Chunks - Organizaciones
- Chunks - Personas
- Chunks - Lugares
- Chunks - Verbos
- Organizaciones - Organizaciones
- Organizaciones - Lugares
- Organizaciones - Verbos
- Personas - Organizaciones
- Personas - Personas

- Personas - Verbos

Un ejemplo de la presentación de los resultados de los análisis de estas parejas de nodos se puede encontrar en la imagen **1-1**, dónde se permite evidenciar el resultado con la pareja organización – verbo. Como se muestra, las organizaciones FARC, TES, DAS y Ejército poseen los círculos con mayores diámetros dentro del grafo, dejando en claro cuáles son los más presentes y cuales poseen un mayor número de relaciones.

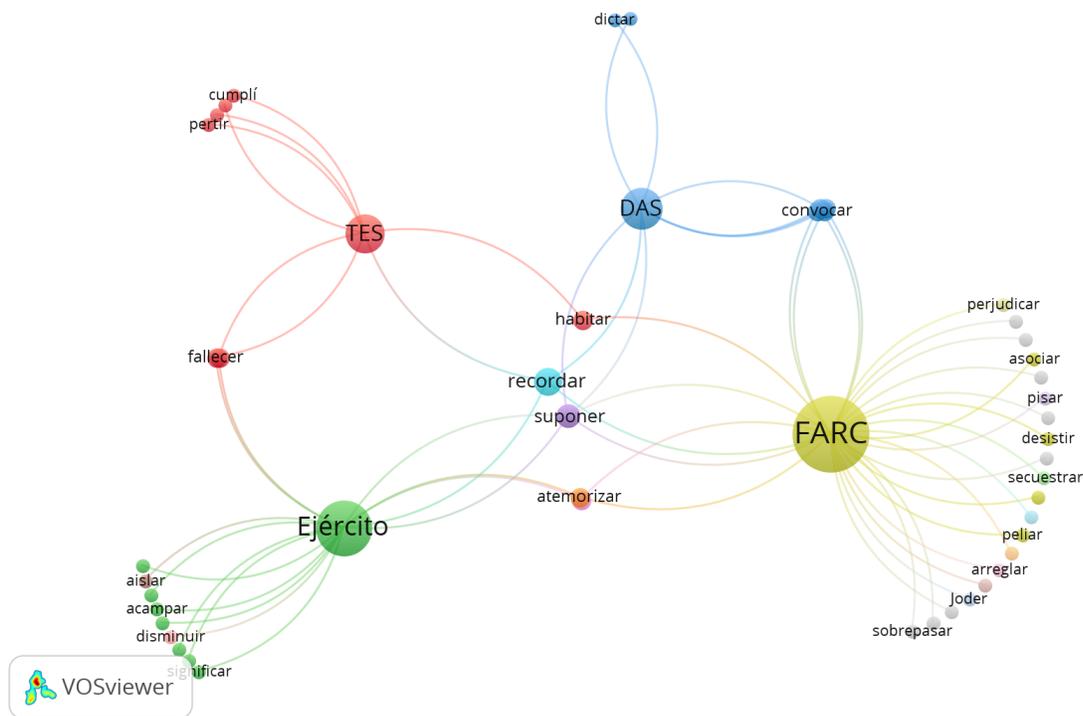


Figura 1-1: Organizaciones-Verbos

Ahora bien, a continuación, se presentarán diferentes ejemplos que apuntan a las tablas desarrolladas de acuerdo con las cuatro métricas estructurales explicadas anteriormente. Un ejemplo de un análisis mediante la métrica estructural All degree se puede encontrar en la tabla **1-1**, dónde se pueden apreciar los ítems más importantes dentro del corpus con sus respectivas frecuencias. Tal y como se muestra, dentro de la pareja de nodos entre organización - verbo, existe una gran relevancia de organizaciones como FARC, el Ejército, el DAS y el TES. Del mismo modo, hay una considerable presencia de verbos como Dar, Saber y Recordar.

Tabla 1-1: All Degree: Organizaciones - Verbos

Rank	Vertex	Value	Id
1	46	30.0000	FARC
2	34	16.0000	Ejército
3	1	9.0000	DAS
4	11	8.0000	TES
5	28	7.0000	dar
6	20	7.0000	saber
7	2	4.0000	recordar
8	6	3.0000	pedir
9	5	3.0000	suponer
10	15	2.0000	fallecer
11	7	2.0000	meter
12	13	2.0000	habitar
13	23	2.0000	comando jungla
14	40	2.0000	abrir
15	9	2.0000	escuchar
16	18	2.0000	presionar
17	36	2.0000	atemorizar
18	4	2.0000	convocar
19	8	2.0000	tocar
20	48	1.0000	Joder

Por otro lado, un ejemplo de la métrica estructural Betweenness centrality se puede encontrar en la tabla **1-2**, dónde se pueden apreciar los ítems con mayor relevancia junto con un valor que, entre más alto se encuentre, mayor será su consideración como un nodo central dentro del corpus. Nuevamente, aparecen en la cima de la tabla las organizaciones como FARC, Ejército y TES, con la inclusión de verbos como Pedir o Suponer.

Tabla 1-2: Betweenness centrality: Organizaciones - Verbos

Rank	Vertex	Value	Id
1	46	0.4249	FARC
2	34	0.2270	Ejército
3	11	0.1028	TES
4	2	0.0961	recordar
5	1	0.0641	DAS
6	6	0.0398	pedir
7	5	0.0398	suponer
8	13	0.0334	habitar
9	28	0.0303	dar
10	20	0.0303	saber
11	40	0.0264	abrir
12	36	0.0264	atemorizar
13	23	0.0236	comando jungla
14	15	0.0104	fallecer
15	18	0.0104	presionar
16	7	0.0048	meter
17	9	0.0048	escuchar
18	4	0.0048	convocar
19	8	0.0048	tocar
20	48	0.0000	Joder

Un ejemplo de la métrica estructural All closeness centrality se puede encontrar en la tabla **1-3**, dónde se pueden apreciar los ítems con mayor lejanía junto con su valor de longitud de distancia. Entre más alto sea el valor, más central es el ítem. Dentro de los ítems con mayores valores se encuentran FARC, recordar, pedir y suponer.

Tabla 1-3: All closeness centrality: Organizaciones - Verbos

Rank	Vertex	Value	Id
1	46	0.4441	FARC
2	2	0.4025	recordar
3	6	0.3645	pedir
4	5	0.3645	suponer
5	40	0.3450	abrir
6	36	0.3450	atemorizar
7	34	0.3360	Ejército
8	13	0.3167	habitar
9	1	0.2995	DAS
10	7	0.2972	meter
11	4	0.2972	convocar
12	9	0.2972	escuchar
13	8	0.2972	tocar
14	11	0.2949	TES
15	60	0.2841	ser
16	52	0.2841	pisar
17	63	0.2841	manecí
18	62	0.2841	operar
19	61	0.2841	generar
20	48	0.2841	Joder

Por último, se tiene el ejemplo de las métricas estructurales Hubs y Authorities. La tabla **1-4** corresponde al Hubs, la cual muestra las organizaciones y verbos con mayor circulación dentro del corpus. Para este caso se obtienen cuatro datos: Dictar, DAS, Recordar y Meter. Por otro lado, la tabla **1-5** representa los resultados de la métrica estructural Authorities, dónde se presentan las organizaciones y verbos con mayor circulación dentro de la red. Para este caso las organizaciones más reiteradas son FARC, Ejército, DAS y TES, mientras que los verbos más circulantes son Recordar, Pedir y Suponer.

Tabla 1-4: Hubs: Organizaciones - Verbos

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	53	80.3030	53	80.3030	dictar
1	3	4.5455	56	84.8485	DAS
2	7	10.6061	63	95.4545	recordar
3	3	4.5455	66	100.0000	meter

Tabla 1-5: Authorities: Organizaciones - Verbos

Rank	Vertex	Value	Id
1	46	0.8669	FARC
2	34	0.3017	Ejército
3	1	0.2805	DAS
4	11	0.1112	TES
5	2	0.0711	recordar
6	6	0.0660	pedir
7	5	0.0660	suponer
8	40	0.0532	abrir
9	36	0.0532	atemorizar
10	7	0.0523	meter

Por lo tanto, gracias a la información suministrada por la imagen **1-1**, se ven las conexiones existentes entre entidades como las FARC con verbos que denotan sucesos influyentes o específicos dentro de las entrevistas asociadas al etiquetado de masacres. Se evidencian las relaciones entre las Fuerzas Armadas como el Ejército Nacional con verbos como Aislar, Acampar y Disminuir, y a las FARC asociadas a verbos como Secuestrar, Desistir, Pisar, Perjudicar, Pelear, etc. Verbos que en este corpus en especial podrían hacer referencia directa a hechos victimizantes cometidos antes, durante o después de una masacre.

Como se explica en el informe de Masacres, en el corpus de entrevistas se encontró que “según los testimonios recolectados, en algunos casos cuando hubo masacre a su vez hubo también tortura, desaparición, amenaza, ataque a bien protegido, secuestro, pillaje, atentado, entre otros. Esta dinámica ocurre porque en ocasiones la persona justo antes o después de ser víctima de masacre puede ser víctima también de otra violación” [2, p. 21]¹. Lo anterior se describe en el apartado g. Elementos característicos de las violaciones/infracciones de dicho informe.

Adicionalmente, el DAS o Departamento Administrativo de Seguridad posee altas frecuencias y se encuentra ligado a verbos como Convocar y Dictar. Sin embargo, a la hora de aplicar las métricas correspondientes se evidencia en las Tablas **1-1**, **1-2**, **1-3**, **1-4** y **1-5** que los verbos más significativos son Dar, Saber, Pedir, Recordar, Abrir y Meter. Todo esto haciendo una mención especial al verbo fallecer que implica “morir (llegar al término de la vida)” [3], y por lo tanto se refiere directamente a un hecho victimizante sufrido por una víctima de masacres.

Como conclusiones generales de este informe, se presentan grandes conexiones entre las FARC y otras organizaciones, lugares o verbos que evidencia la correspondencia de esta or-

¹Versión aún en revisión

ganización con actos relacionados con masacres per se. Del mismo modo, se generan verbos relacionados con acciones violentas dentro del componente de masacres por parte de diversos actores. Lo cual, deja en evidencia el entramado complejo que compone a organizaciones y actores partícipes del conflicto armado, específicamente con el componente de la masacre.

Otro aspecto importante para concluir es la delimitación geográfica de las masacres, o, por lo menos, de las relaciones presentes dentro del corpus bajo esta etiqueta. El espacio geográfico es ocupado por Antioquia, puesto que el departamento colombiano, además de estar bastante presente dentro del corpus, posea una amplia circulación producto de sus numerables relaciones con personas y organizaciones. No obstante, todos los resultados permiten analizar de manera eficaz y eficiente los diferentes datos procesados. Es importante reconocer que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas. Incluso en esta afirmación es importante tener cuidado, puesto que algunas expresiones dialécticas recogidas en forma de sintagma nominal, pueden ayudar a entender mejor la relación entre nodos. Un ejemplo de esto es el uso del término “tas tas” que ayuda a ligarlo con diferentes organizaciones y, en especial, con el verbo matar.

1.2. Análisis de redes de relaciones presentes en el corpus de entrevistas de Justicia y Paz

Este informe fue culminado el 8 de febrero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. En el presente informe se detalla el trabajo de procesamiento de corpus de las entrevistas recolectadas por la Comisión de la Verdad relacionadas con el tema de Justicia y Paz. Dentro de las mismas, se busca encontrar patrones y conexiones representadas por sistemas de redes que permitan generar un estudio a mayor profundidad dentro de Justicia y Paz. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Organizaciones - Organizaciones
- Organizaciones - Lugares
- Organizaciones - Verbos
- Personas - Organizaciones
- Personas - Personas
- Personas - Verbos

La esquematización de los resultados de esta sección obedece a las mismas leyes de la anterior sección y, del mismo modo, se presentarán las mismas ejemplificaciones analíticas para las demás secciones que compartan la misma metodología y presentación de resultados. Un ejemplo de la presentación de los resultados de los análisis de estas parejas de nodos se puede encontrar en la imagen **1-2**, dónde se permite evidenciar el análisis gráfico de la pareja de nodos organizaciones – verbos. Como se presenta, los verbos más frecuentes y con mayores relaciones se vinculan directamente con dinámicas del conflicto armado, específicamente aspectos violentos: disparar, hacer o pertenecer. Del mismo modo, salta a la vista la asociación de estos verbos con organizaciones como el Frente Barro Blanco, la Fiscalía, el Ejército o las FARC.

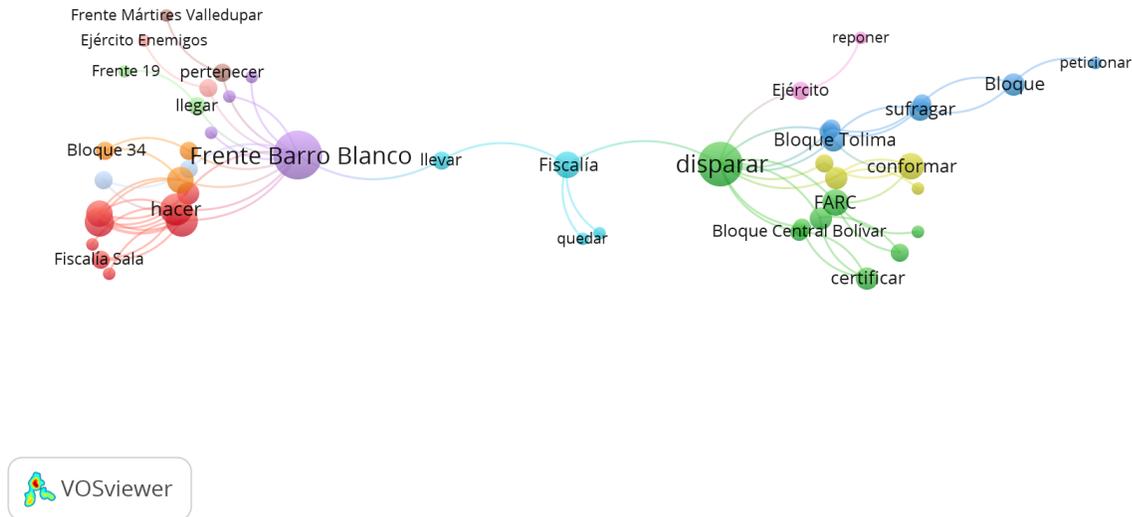


Figura 1-2: Organizaciones-Verbos

Por lo tanto, gracias a la información suministrada en la imagen **1-2**, se presentan las conexiones entre las diferentes organizaciones y los verbos que denotan sucesos influyentes o específicos dentro del corpus. La alta presencia de la Fiscalía se explica por las negociaciones de paz realizadas durante el primer periodo presidencial de Álvaro Uribe Vélez (2002-2006) entre el gobierno nacional y los grupos paramilitares. En 2005, se aprueba la ley 975 con el fin de: “facilitar los procesos de paz y la reincorporación individual o colectiva a la vida civil de miembros de grupos armados al margen de la ley, garantizando los derechos de las víctimas a la verdad, la justicia y la reparación” [4]. La Fiscalía se encargaría de varias funciones dentro

de lo contemplado por la ley, como por ejemplo el otorgamiento de los beneficios que acobija la ley de acuerdo al listado entregado por parte del Gobierno Nacional, la conservación de los registros por parte del Secretario de la Unidad Nacional de Fiscalía para la Justicia y Paz, la investigación de las circunstancias de tiempo, modo y lugar en que se realizaron las conductas punibles, la velación de la protección de las víctimas, testigos y los peritos que pretenda presentar en juicio, entre otros [4].

Ahora bien, la presencia de las organizaciones de los grupos ilegales se explica, en su relación con los verbos presentes, por los procesos de desmovilización y acobijación dentro de la ley 975 de 2005. Un ejemplo de esto sucede con la organización del Frente Barro Blanco, el cual es una facción del gran Bloque Mineros de las AUC, el cual se desmoviliza el 20 de enero de 2006. [5].

Ahora bien, para el caso de la métrica estructural All Degree, se puede destacar el ejemplo ilustrado de la tabla **1-6**, donde se presentan los términos más frecuentes de la pareja de nodos organización – organización. Gracias al análisis de la tabla, se permite entender que las organizaciones más importantes dentro del componente de Justicia y Paz hacen parte de la Fiscalía, las AUC y el Bloque Norte.

Tabla 1-6: All Degree: Organización - Organización

Rank	Vertex	Value	Id
1	20	16.0000	Fiscalía
2	33	12.0000	AUC
3	12	9.0000	Bloque Norte
4	2	7.0000	Fiscalía General Nación
5	6	5.0000	Autodefensas
6	8	4.0000	FARC
7	7	3.0000	Cuarta Brigada
8	10	3.0000	Fuerzas Armadas
9	31	2.0000	Bloque Capital
10	18	2.0000	ANI
11	1	2.0000	Frente Barro Blanco
12	28	1.0000	UNJP
13	27	1.0000	Frente 22 FARC
14	26	1.0000	Fiscalía 6
15	25	1.0000	Bloque Autodefensas Campesinas Cundinamarca
16	24	1.0000	Fiscalía Once
17	23	1.0000	Administración justicia
18	22	1.0000	Secretaría Transito Valledupar
19	21	1.0000	Procuraduría General Nación
20	5	1.0000	oficina Envigado

Dentro del ejemplo de la métrica estructural de Betweenness centrality recogida bajo el ejemplo de la tabla **1-7**, se puede observar que los nodos centrales adjudicados en el tratamiento de datos de la pareja de nodos organización – organización son Fiscalía, Bloque Norte y Autodefensas. Como se puede apreciar, los mismos nodos centrales de esta pareja de nodos representan también los términos más frecuentes de la tabla **1-6**.

Tabla 1-7: Betweenness centrality: Organización - Organización

Rank	Vertex	Value	Id
1	20	0.5363	Fiscalía
2	12	0.3514	Bloque Norte
3	6	0.3479	Autodefensas
4	33	0.3351	AUC
5	2	0.2058	Fiscalía General Nación
6	8	0.0616	FARC
7	18	0.0501	ANI
8	7	0.0451	Cuarta Brigada
9	10	0.0451	Fuerzas Armadas
10	31	0.0246	Bloque Capital
11	29	0.0000	WAC
12	28	0.0000	UNJP
13	27	0.0000	Frente 22 FARC
14	26	0.0000	Fiscalía 6
15	25	0.0000	Bloque Autodefensas Campesinas Cundinamarca
16	24	0.0000	Fiscalía Once
17	23	0.0000	Administración justicia
18	22	0.0000	Secretaría Transito Valledupar
19	21	0.0000	Procuraduría General Nación
20	5	0.0000	oficina Envigado

Para el análisis de la tercera métrica estructural llamada All closeness centrality, se debe evaluar la tabla **1-8**, la cual se compone de los nodos más lejanos y cercanos de la pareja de nodos organizaciones – organizaciones. Dentro de los resultados se pueden apreciar los términos con mayores valores, los cuales se traducen en los más cercanos entre si como lo son Autodefensas y la Fiscalía. Se mantiene presente la congruencia entre los términos más frecuentes y sus relaciones dentro del corpus de Justicia y Paz.

Tabla 1-8: All closeness centrality: Organización - Organización

Rank	Vertex	Value	Id
1	6	0.5362	Autodefensas
2	20	0.5139	Fiscalía
3	2	0.4933	Fiscalía General Nación
4	33	0.4568	AUC
5	7	0.4157	Cuarta Brigada
6	12	0.4157	Bloque Norte
7	10	0.4157	Fuerzas Armadas
8	31	0.3978	Bloque Capital
9	8	0.3737	FARC
10	1	0.3663	Frente Barro Blanco
11	18	0.3663	ANI
12	30	0.3426	Escuela Américas 42
13	29	0.3426	WAC
14	28	0.3426	UNJP
15	27	0.3426	Frente 22 FARC
16	26	0.3426	Fiscalía 6
17	25	0.3426	Bloque Autodefensas Campesinas Cundinamarca
18	24	0.3426	Fiscalía Once
19	23	0.3426	Administración justicia
20	22	0.3426	Secretaría Transito Valledupar

Por último, las tablas que ayudan a ejemplificar el uso de las métricas estructurales de Hubs y Authorities se encuentran a continuación. La tabla **1-9** representa las entidades u organizaciones con mayor circulación en el corpus, arrojando los resultados de Autodefensas bloque resistencia Tayrona y Frente Barro Blanco. De esta manera, se evidencia que no necesariamente todas las organizaciones que poseen una alta relevancia dentro del corpus circulan de manera congruente. Por otra parte, la tabla **1-10** representa las entidades con mayor circulación dentro de la red. De esta manera, los valores que más circulan son Fiscalía y las AUC, entidades que corresponden de igual manera a los análisis de las demás métricas estructurales.

Tabla 1-9: Hubs: Organización - Organización

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	28	73.6842	28	73.6842	Autodefensas bloque resistencia Tayrona
2	10	26.3158	38	100.0000	Frente Barro Blanco

Tabla 1-10: Authorities: Organización - Organización

Rank	Vertex	Value	Id
1	20	0.4989	Fiscalía
2	33	0.4041	AUC
3	2	0.3257	Fiscalía General Nación
4	6	0.3147	Autodefensas
5	7	0.2211	Cuarta Brigada
6	10	0.2211	Fuerzas Armadas
7	31	0.1871	Bloque Capital
8	8	0.1639	FARC
9	1	0.1512	Frente Barro Blanco
10	12	0.1258	Bloque Norte

Gracias a la información suministrada en las anteriores tablas, es posible observar un alto valor del Bloque Norte (junto con la Fiscalía y las AUC). Esto puede deberse a los diálogos adelantados entre este grupo, perteneciente a las Autodefensas Unidas de Colombia, y el Gobierno Nacional. Estos diálogos se presentaron previos a la vigencia de la ley 975 de 2005 y, por lo tanto, entra a debate en la claridad de los actos de desmovilización, diálogos y acuerdos de paz dentro de la ley 975. Esto se debe a que el artículo 72 de esta ley definía que: “La presente ley deroga todas las disposiciones que le resulten contrarias. Se aplicará únicamente a hechos ocurridos con anterioridad a su vigencia y rige a partir de la fecha de su promulgación” [4]. La necesidad de la claridad se encontraba en que, las negociaciones de una desmovilización de miembros por un proceso de paz del Bloque Norte de las AUC no podían ser derogadas por el artículo 72, aunque tampoco podían ser acobijados por el beneficio de alternatividad por conductas punibles cometidas por el grupo después del 25 de julio de 2005 [4].

Por último, en la tabla 1-8, en la cual se presentan las organizaciones con mayor lejanía dentro del corpus, se pueden destacar la Secretaría de Tránsito de Valledupar y la Administración de Justicia. La explicación de esta lejanía en cuanto a la secretaría de Tránsito puede deberse a que no es una organización relacionada directamente con el tema de Justicia y Paz, no obstante, es importante mencionar que, en Valledupar, de conformidad con lo dispuesto en el inciso segundo del artículo 53 de la ley 975 de 2005, operó la Comisión Regional para la Restitución de Bienes [4]. A su vez, la Administración de Justicia fue el espacio al cual podían acceder las víctimas dónde recibirían un trato humano digno, protección de intimidad y seguridad personal y familiar, pronta reparación de los daños sufridos, serían oídas y facilitadas con el aporte de pruebas, información pertinente para la protección de sus intereses y conocer la verdad, ser asistidas durante el juicio por un abogado, recibir asistencia integral para su recuperación y a ser asistidas gratuitamente por un traductor o intérprete [4].

Como conclusiones generales de este informe, es importante resaltar la relevancia e importancia de la Fiscalía como una de las organizaciones considerada como un nodo central dentro del análisis de resultados. Del mismo modo, esto solo es posible por la alta presencia de los grupos paramilitares y de autodefensas que se encuentran bastante frecuentes. Esto indica que el proceso desarrollado y enmarcado alrededor de justicia y paz mantiene una estrecha relación entre los oficios de la fiscalía y los acontecimientos y espacios de diálogo proporcionados por integrantes de grupos de autodefensas. Esta relación también adquiere fuerza debido a la amplia participación dentro del corpus y de la red de espacios geográficos como Cundinamarca.

Como conclusión adicional, es importante un estudio a mayor profundidad acerca del uso de los verbos dentro de este informe en particular. Esto se debe a que gran parte de las relaciones entre personas y organizaciones con los verbos suscitan a acciones violentas. Esto se sustenta en el hecho de que verbos como disparar y encontrar se encuentren tan relevantes al interior del corpus y de la red.

Por último, es importante recalcar, al igual que lo dicho en la primera sección de este capítulo, que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas.

1.3. Análisis de redes de entrevistas de dinámicas del conflicto

Este informe fue culminado el 26 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. En el presente informe se detalla el trabajo de procesamiento de corpus de las entrevistas recolectadas por la Comisión de la Verdad, a través de las cuales se busca responder la pregunta de ¿quién le hizo qué a quién? Dentro de las mismas, se busca encontrar patrones y conexiones representadas por sistemas de redes y procesadas según reportes de métricas estructurales que ayuden a analizar y comprender las dinámicas del conflicto armado colombiano. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Chunks - Adjetivos
- Chunks - Organizaciones
- Chunks - Personas
- Chunks - Lugares

- Chunks - Verbos
- Organizaciones - Organizaciones
- Organizaciones - Lugares
- Organizaciones - Verbos
- Personas - Organizaciones
- Personas - Personas
- Personas - Verbos

Un ejemplo de la presentación de los resultados de los análisis de estas parejas de nodos se puede encontrar en la imagen **1-3**, dónde se permite evidenciar el análisis gráfico de la pareja de nodos organizaciones – verbos. De acuerdo con los resultados, los verbos más frecuentes, junto con sus relaciones con algunas instituciones nacionales, permiten entrever, de cierta manera, los percances y objetivos en el marco del conflicto armado. Esto se debe a que son verbos como sacar o alcanzar. Las instituciones relacionadas son algunas como el ejército, la personería y la comisión.

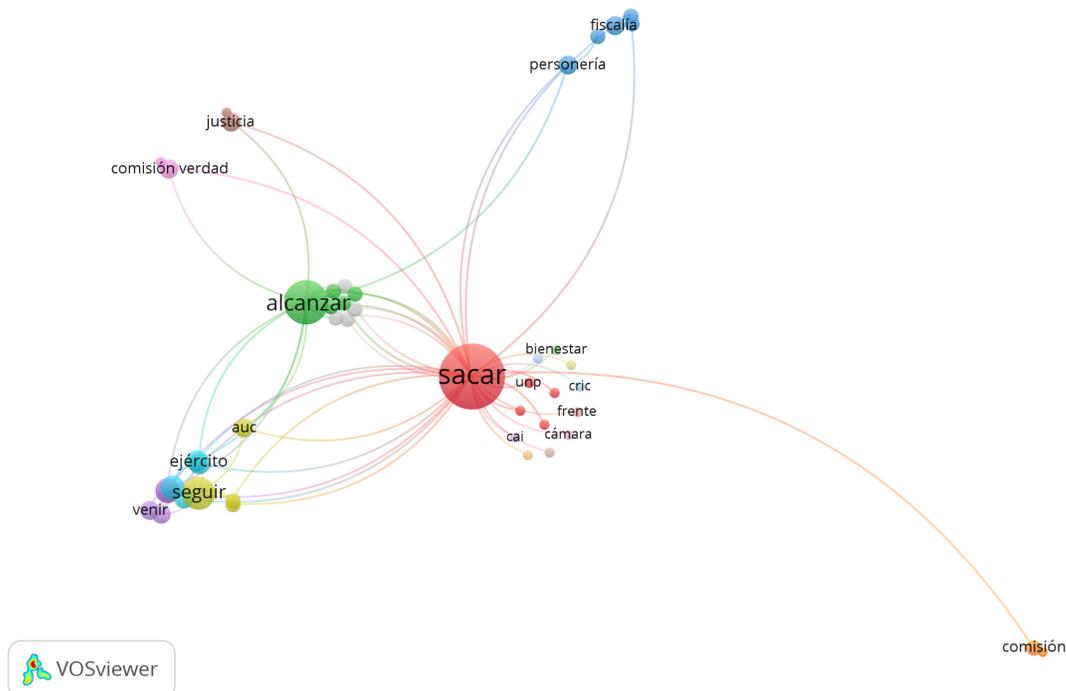


Figura 1-3: Organizaciones-Verbos

Dentro de la primera métrica estructural, All Degree, se puede desarrollar un análisis a partir del ejemplo de la tabla **1-11**, la cual presenta los términos más frecuentes de la pareja de nodos organización – verbos. Gracias al análisis de la tabla, se permite entender que las organizaciones más importantes en la búsqueda por establecer el agente y el destinatario del conflicto armado son el ELN y el M-19. Por otro lado, los verbos más preponderantes son sacar, alcanzar y seguir.

Tabla 1-11: All Degree: Organización - Verbo

Rank	Vertex	Value	Id
1	1	36.0000	sacar
2	18	16.0000	alcanzar
3	20	9.0000	seguir
4	46	5.0000	ELN
5	45	5.0000	M19
6	39	5.0000	permitir
7	44	4.0000	policía
8	43	4.0000	ejército
9	31	3.0000	auc
10	42	3.0000	personería
11	40	3.0000	comisión verdad
12	38	3.0000	cti
13	36	3.0000	justicia
14	34	3.0000	fiscalía
15	33	3.0000	venir
16	32	3.0000	procuraduría
17	15	2.0000	comisión
18	30	2.0000	unión patriótica
19	29	2.0000	autodefensas
20	28	2.0000	junta

En la métrica estructural del Betweenness centrality encontrada en la tabla **1-12**, se pueden denotar los términos denominados nodos centrales al conseguir un puntaje mayor. Entre estos términos se encuentran los verbos sacar y alcanzar, mientras que las organizaciones más frecuentes son justicia y comisión. No obstante, al no tener un complemento para justicia, es complicado asociarlo con una sociedad exacta, más allá del determinismo político de la corte de justicia, por ejemplo.

Tabla 1-12: Betweenness centrality: Organización - Verbo

Rank	Vertex	Value	Id
1	1	0.8257	sacar
2	18	0.0972	alcanzar
3	36	0.0458	justicia
4	40	0.0458	comisión verdad
5	15	0.0444	comisión
6	34	0.0400	fiscalía
7	46	0.0299	ELN
8	45	0.0299	M19
9	42	0.0263	personería
10	26	0.0210	sena
11	20	0.0200	seguir
12	32	0.0144	procuraduría
13	44	0.0142	policía
14	43	0.0142	ejército
15	38	0.0095	cti
16	31	0.0059	auc
17	39	0.0036	permitir
18	19	0.0032	justicia paz
19	21	0.0032	defensoría pueblo
20	30	0.0014	unión patriótica

Ahora bien, en relación con la métrica estructural del All closeness centrality, es interesante analizarla mediante el ejemplo de la tabla **1-13**. En esta ocasión, se puede evidenciar como los términos sacar, ELN y M-19 poseen una centralidad muy alta, con muy poco distanciamiento. No obstante, es importante añadir que, aunque estos tres términos sean centrales, el verbo sacar posee una centralidad muy superior al segundo término, resaltándolo sobre todos los demás.

Tabla 1-13: All closeness centrality: Organización - Verbo

Rank	Vertex	Value	Id
1	1	0.8333	sacar
2	46	0.5000	ELN
3	45	0.5000	M19
4	44	0.4891	policía
5	43	0.4891	ejército
6	31	0.4787	auc
7	42	0.4787	personería
8	40	0.4787	comisión verdad
9	38	0.4787	cti
10	36	0.4787	justicia
11	34	0.4787	fiscalía
12	32	0.4787	procuraduría
13	15	0.4688	comisión
14	30	0.4688	unión patriótica
15	29	0.4688	autodefensas
16	28	0.4688	junta
17	26	0.4688	sena
18	25	0.4688	acción social
19	24	0.4688	up
20	23	0.4688	JEP

Por último, se tienen las ejemplificaciones de las métricas estructurales de Hubs y Authorities. Para el primer caso, se encuentra la tabla **1-14**, la cual presenta la circulación de los términos dentro de la colección. Dentro de los términos más circulados se encuentran frente, alcanzar, sacar y DAS. Este caso es bastante particular, puesto que el verbo sacar continúa demostrando preponderancia, esto es algo que no sucede de manera tan evidente en las secciones anteriores. De la misma forma, para el segundo caso, se encuentra la tabla **1-15**, la cual presenta los términos que más circulan dentro de la red. Para este caso dichos términos son sacar, alcanzar, seguir y permitir. Nuevamente, la constancia del verbo sacar se presenta. Otro aspecto curioso por analizar es que ninguna organización consigue circular dentro de la red de la misma manera que los verbos.

Tabla 1-14: Hubs: Organización - Verbo

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	32	69.5652	32	69.5652	frente
1	4	8.6957	36	78.2609	alcanzar
2	6	13.0435	42	91.3043	sacar
3	4	8.6957	46	100.0000	DAS

Tabla 1-15: Authorities: Organización - Verbo

Rank	Vertex	Value	Id
1	1	0.7776	sacar
2	18	0.4357	alcanzar
3	20	0.2527	seguir
4	39	0.1589	permitir
5	33	0.0920	venir
6	46	0.0828	ELN
7	45	0.0828	M19
8	44	0.0784	policía
9	43	0.0784	ejército
10	31	0.0707	auc

Retomando la importancia del establecimiento de relaciones de las organizaciones relevantes del corpus con los verbos portadores de acción y sentido dentro del corpus, dentro de estos resultados se evidencian verbos como Alcanzar o Sacar. Para el caso del primero, es definido como “Llegar a juntarse con alguien o algo que va delante” [3](definición 1) y “Llegar a tocar, golpear o herir a alguien o algo” [3](definición 2), en un contexto más violento. Igualmente, el verbo cuenta con otras definiciones que lo acercan a Percibir, Notar, Dar o Saber. La relevancia de este verbo dentro del corpus podría deberse no solo a la ambigüedad del mismo, sino al papel del testigo en alcanzar a ver, conocer o saber algo o al de la víctima de recibir una acción violenta en determinado tiempo o que es concluida. Hechos como los perpetrados por diversos actores, según el informe, “en el primer clúster asociado al accionar de la Fuerza Pública, los Terceros y Otros Actores Armado se concentran 96 de cada 100 homicidios (96.1 %; n=6.036), 93 de cada 100 detenciones (93.3 %; n=808), 90 de cada 100 desapariciones (90.6 %; n=1.807), 88 de cada 100 secuestros (88.3 %, n=1.429), más de la mitad de los exilios (57.3 %, n=1.027) y de las torturas físicas (55.9 %; n=1.502)” [2, p. 3].

Para el caso del segundo verbo, este es definido como “Poner algo fuera del lugar donde estaba encerrado o contenido” [3] (Definición 1). Este último podría estar ligado a los desplazamientos forzados o actos violentos en los que “se saca” a las personas de su hogar o

de su lugar de residencia. Hechos que se referencian en el informe, analizados tanto como hechos victimizantes catalogados por actores como a manera de impactos individuales donde se “ocasionaron daños de discapacidad sensorial y física (41.8%) y de desplazamiento (36.5%)” [2, p. 7].

Adicionalmente, las Tablas **1-11**, **1-12**, **1-13**, **1-14** y **1-15** muestran la relevancia de organizaciones como el ELN, el M-19, la Policía junto a los verbos Sacar, Alcanzar y Seguir discutido anteriormente. Además se encuentra un alto valor en el verbo Permitir y las Autodefensas Unidas de Colombia. El primero implica una posición de poder de quien ejerce la acción sobre quien la recibe y la segunda una entidad reconocida como perpetradora de hechos victimizantes en el marco del conflicto armado.

Como conclusiones generales, es importante analizar como las FARC y la Comisión se encuentran tan presentes en los diversos casos de análisis de pares de nodos, teniendo en cuenta que esto significa que estas organizaciones poseen una alta frecuencia dentro del corpus, una alta circulación dentro de este mismo y una alta circulación dentro de la red. De la misma manera, es interesante ver como se conecta con verbos como sacar o alcanzar, los cuales terminan siendo protagonistas en todos los análisis concernientes con esta categoría. Esto se podría relacionar con la capacidad de estas organizaciones de alcanzar ciertos tipos de objetivos, ya sea metafísicos o naturales. A su vez, es interesante analizar estos resultados verbales, puesto que se dan a entender como verbos con un propósito más auxiliar que un tipo de violencia claro. Tanto sacar como alcanzar auxilian en la funcionalidad de objetivos por parte de las organizaciones.

Ahora bien, otro aspecto interesante a resaltar es la delimitación geográfica del análisis de estas entrevistas. Esto arroja una gran constancia entre la capital del país y diversas organizaciones, personas y sintagmas nominales. Permittiendo entender, así, que Bogotá se convierte en un punto referente a la hora de gestionar un análisis con un contenido importante de lugar de producción.

Por último, es importante recalcar, al igual que lo dicho en las anteriores secciones de este capítulo, que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas.

1.4. Análisis de redes de entidades presentes en entrevistas etiquetadas por el equipo de resistencia

Este informe fue culminado el 26 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. En este informe se tiene como interés principal presentar un análisis basado en el corpus compuesto por las entrevistas determinadas por el equipo de resistencia. Esto con el objetivo de presentar un manejo y análisis de datos que haga posible la determinación de bases para futuros procesos de investigación, el esclarecimiento y el entendimiento de lo sucedido en el marco del conflicto armado en Colombia. De acuerdo con la metodología aplicada, las parejas de nodos desarrollados para analizar los datos del corpus fueron los siguientes:

- Chunks - Organizaciones

- Chunks - Lugares

- Chunks - Verbos

- Organizaciones - Organizaciones

- Organizaciones - Lugares

- Organizaciones - Verbos

Un ejemplo de la presentación de los resultados de los análisis de estas parejas de nodos se puede encontrar en la imagen **1-4**, dónde se permite evidenciar el análisis gráfico de la pareja de nodos chunks - lugares. De acuerdo con los resultados, se pueden establecer relaciones claras de delimitaciones geográficas de procesos y acontecimientos desarrollados a nivel nacional, regional o local. Para el caso de la imagen **1-4**, estos procesos se encuentran presentes a nivel local, en específico, con las ciudades de Medellín y Bogotá. Dentro de los sintagmas nominales más frecuentes, estos son más complicados de delimitar, puesto que, como se permite evidenciar en la imagen, la gran mayoría de términos poseen un tamaño de referencia circunferencial bastante similar.

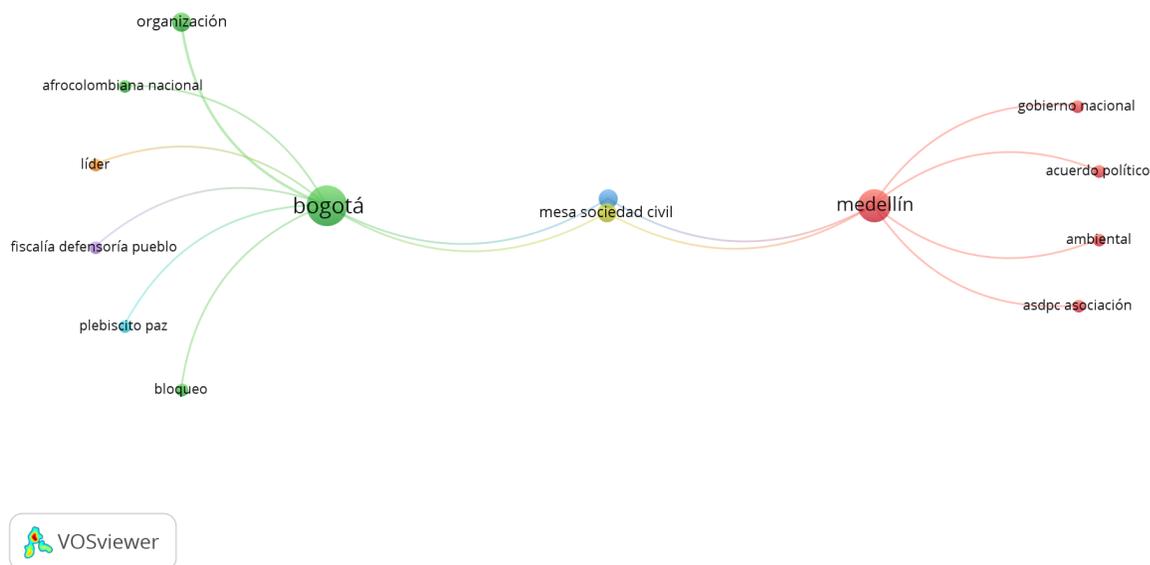


Figura 1-4: Chunks - Lugares

A continuación, se procede al análisis de las diferentes métricas estructurales desarrolladas dentro de la metodología de estudio. Para el caso de la primera métrica, All Degree, se presenta la tabla **1-16**, la cual considera la pareja de nodos chunks – lugares. La tabla contiene los términos más frecuentes o relevantes dentro del corpus. Entre estos se encuentran los lugares Bogotá y Medellín, mientras que por parte de los chunks o sintagmas nominales se encuentra marcha patriótica. Esta información coincide con lo visto en el imagen **1-4** de la misma pareja de nodos, puesto que los lugares resaltan demasiado y los sintagmas nominales son bastante parecidos unos con otros. Aunque se haya mencionado la marcha patriótica en específico, como se puede ver en la tabla, la gran mayoría de chunks poseen un valor bastante similar.

Tabla 1-16: All Degree: Chunks - Lugares

Rank	Vertex	Value	Id
1	8	9.0000	bogotá
2	2	6.0000	medellín
3	3	2.0000	marcha patriótica
4	11	2.0000	organización
5	1	2.0000	mesa sociedad civil
6	10	1.0000	fiscalía defensoría pueblo
7	9	1.0000	bloqueo
8	4	1.0000	asdpc asociación
9	7	1.0000	gobierno nacional
10	6	1.0000	ambiental
11	14	1.0000	líder
12	12	1.0000	plebiscito paz
13	13	1.0000	afrocolombiana nacional
14	5	1.0000	acuerdo político

Para analizar la segunda métrica estructural, Betweenness centrality, se procede a observar la ejemplificación desarrollada en la tabla **1-17**. En esta, se pueden observar los considerados nodos centrales de acuerdo con un alto valor en la columna Value. Es evidente, en consecuencia, que los nodos centrales son Bogotá y Medellín. En relación con los sintagmas nominales, únicamente la marcha patriótica y la mesa de sociedad civil poseen datos, el resto se encuentran en 0 al ser poco relevantes. De esta manera se continúa la congruencia mostrada previamente.

Tabla 1-17: Betweenness Centrality: Chunks - Lugares

Rank	Vertex	Value	Id
1	8	0.7372	bogotá
2	2	0.5449	medellín
3	3	0.2244	marcha patriótica
4	1	0.2244	mesa sociedad civil
5	11	0.0000	organización
6	10	0.0000	fiscalía defensoría pueblo
7	9	0.0000	bloqueo
8	4	0.0000	asdpc asociación
9	7	0.0000	gobierno nacional
10	6	0.0000	ambiental
11	5	0.0000	acuerdo político
12	14	0.0000	líder
13	13	0.0000	afrocolombiana nacional
14	12	0.0000	plebiscito paz

All closeness centrality, perteneciente a la tercera red métrica estructural, se puede encontrar representada y ejemplificada en la tabla **1-18**. En esta ocasión, se presentan los términos más cercanos y alejados dentro del corpus. Estos términos son Bogotá, marcha patriótica, mesa sociedad civil y Medellín. Como se puede apreciar, en este caso los sintagmas nominales poseen un alto valor dentro de estos datos, por lo tanto, aunque no se encuentren tan presentes o relevantes, si poseen una cercanía entre los distintos términos del corpus.

Tabla 1-18: All Closeness Centrality: Chunks - Lugares

Rank	Vertex	Value	Id
1	8	0.5909	bogotá
2	3	0.5417	marcha patriótica
3	1	0.5417	mesa sociedad civil
4	2	0.5000	medellín
5	14	0.3824	líder
6	10	0.3824	fiscalía defensoría pueblo
7	13	0.3824	afrocolombiana nacional
8	12	0.3824	plebiscito paz
9	11	0.3824	organización
10	9	0.3824	bloqueo
11	4	0.3421	asdpc asociación
12	7	0.3421	gobierno nacional
13	6	0.3421	ambiental
14	5	0.3421	acuerdo político

Por último, en relación con la cuarta red métrica estructural que ejemplifica el Hubs y Authorities, se encuentran las tablas **1-19** y **1-20** respectivamente. La primera tabla representa los términos que más circulan dentro del corpus, de los cuales únicamente se encuentran asdpc asociación y mesa sociedad civil. De esta forma, se demuestra que los lugares en realidad no se encuentran circulando en bastantes cantidades dentro del corpus, no obstante, algunos chunks, como es el caso del segundo resultado, mantienen su circulación y relevancia en relación con su cercanía con los demás nodos. Ahora bien, la segunda tabla representa los términos que más circulan dentro de la red. En esta ocasión, los lugares más presentes vuelven a ser Bogotá y Medellín, mientras que los chunks más presentes son organización y marcha patriótica. Como se evidenciar, este resultado es más congruente con las demás estructuras métricas, ignorando el Hubs.

Tabla 1-19: Hubs: Chunks - Lugares

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	4	28.5714	4	28.5714	asdpc asociación
2	10	71.4286	14	100.0000	mesa sociedad civil

Tabla 1-20: Authorities: Chunks - Lugares

Rank	Vertex	Value	Id
1	8	0.8712	bogotá
2	2	0.3056	medellín
3	11	0.2120	organización
4	3	0.1431	marcha patriótica
5	1	0.1431	mesa sociedad civil
6	12	0.1060	plebiscito paz
7	14	0.1060	líder
8	13	0.1060	afrocolombiana nacional
9	10	0.1060	fiscalía defensoría pueblo
10	9	0.1060	bloqueo

Por lo tanto, en relación con las tablas y figura de la red que relaciona los chunks y los lugares, hay relevancia en términos que representan lugares como Bogotá y Medellín, y chunks que representan a organizaciones como la Mesa Sociedad Civil y Marcha patriótica. En la imagen 1-4, las tablas 1-18 y 1-19 se evidencia la centralidad del último término mencionado; en la tabla 1-20 esta entidad es presentada como uno de los elementos de mayor circulación dentro del corpus. Otros chunks presentados en los resultados se refieren a movimientos sociales y otros tipos de organizaciones que se establecen principalmente dentro del Estado con propósitos e intereses específicos. Algunos de estos son la Mesa Nacional de Sociedad Civil (mostrado en la tabla 1-18 como un chunk de alta circulación en el corpus).

Finalmente, teniendo en cuenta las relaciones establecidas entre estos dos tipos de términos, estas se caracterizan principalmente en iniciativas, hechos y algunos sujetos sociales como los mostrados en la tabla 1-20 (organización, marcha patriótica) que se desarrollan específicamente en el contexto nacional. Uno de estos ejemplos puede ser el movimiento Marcha Patriótica, que fue fundado en el año 2012 en Colombia y se ha centrado en la reivindicación de "...la construcción de un país soberano y de una paz con justicia social." [6].

Como conclusiones generales, es importante recalcar que las organizaciones en esta investigación no se encuentran tan presentes. Es más, en todos los análisis, excepto el de organización – organización y organización – lugar, los términos más recurrentes, relevantes, cercanos y circulantes dentro de la red son chunks, o lugares. Gracias a esto, se puede concluir que la investigación se encuentra más direccionada a los verbos, sintagmas nominales y los lugares. Entre los verbos más destacados se encuentran tener, hablar o hacer. Estos verbos por si solos no desarrollan un aspecto violento como en secciones pasadas, sino que, dependiendo del complemento, se pueden encontrar aspectos importantes dentro de los lugares, organizaciones o sintagmas nominales que evalúan dichas relaciones.

Del mismo modo, la presencia de ciudades principales del país como Medellín o Bogotá, permiten entender que las dimensiones del corpus corresponden a aspectos más urbanos o, en otra medida, los lugares donde se recogen los testimonios con presencia de la institucionalidad. Esto se debe, a su vez, a que la relación desarrollada con algunos chunks presenta movilizaciones o desarrollos socio-políticos acontecidos en estos espacios.

Por último, es importante recalcar, al igual que lo dicho en las anteriores secciones de este capítulo, que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas.

1.5. Análisis de redes de entidades presentes en sentencias de la Unidad de Restitución de Tierras (URT) Nariño

Este informe fue culminado el 25 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Este documento es el resultado del trabajo analítico del corpus de sentencias de la Unidad de Restitución de Tierras (URT). Si bien las víctimas juegan un papel demasiado importante en este informe, se busca ahondar en las personas u organizaciones que aparecen como victimarios, ya que desde este punto de vista se puede entender el cómo y el porqué del conflicto armado desde otra perspectiva. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron las siguientes:

- Organizaciones - Organizaciones
- Organizaciones - Lugares
- Organizaciones - Verbos
- Personas - Organizaciones
- Personas - Personas
- Personas - Verbos

Un ejemplo de la presentación de los resultados de los análisis de estas parejas de nodos se puede encontrar en la imagen **1-5**, donde se permite evidenciar el análisis gráfico de la pareja de nodos organizaciones - organizaciones. De acuerdo con los resultados, se pueden establecer relaciones entre diferentes entidades de corte legal e ilegal. Para el caso de la imagen, se desarrollan relaciones entre las FARC y el ejército, o las FARC y la policía. No

obstante, este tipo de relaciones son bastante escasas, puesto que la gran concentración se ubica entre grupos ilegales, como es el caso de las Autodefensas Campesinas y las Águilas Negras. Lo anterior se menciona puesto que, además de ser lo más frecuente, también posee los ítems más recurrentes dentro del corpus, como lo son las Autodefensas Campesinas.

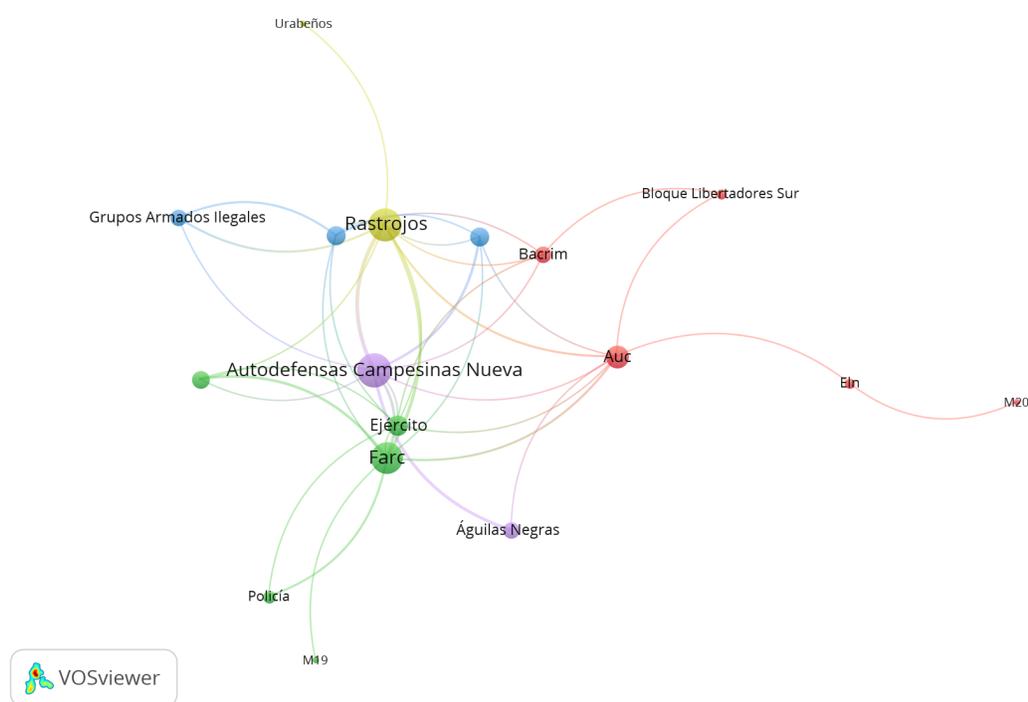


Figura 1-5: Organización - Organización

La primera métrica estructural por analizar corresponde a All Degree. Para ello, se hace uso de la tabla **1-21** como ejemplificación de esta métrica con la pareja de nodos organizaciones – organizaciones, igual que la analizada para el gráfico de la imagen **1-5**. La tabla contiene los términos más frecuentes o relevantes dentro del corpus. Entre estos se encuentran las Autodefensas Campesinas Nueva Generación, los Rastrojos y las FARC. Estas tres organizaciones obtienen valores cercanos, no obstante, en relación con las demás organizaciones, estas se encuentran bastante alejadas del valor.

Tabla 1-21: All Degree: Organización - Organización

Rank	Vertex	Value	Id
1	7	21.0000	Autodefensas Campesinas Nueva Generación
2	6	20.0000	Rastrojos
3	2	18.0000	Farc
4	15	10.0000	Auc
5	4	8.0000	Ejército
6	1	7.0000	Bloque Central Bolívar
7	5	6.0000	Águilas Negras
8	18	6.0000	Fuerza Pública
9	9	5.0000	Bacrim
10	16	5.0000	Águilas Negras
11	8	4.0000	Grupos Armados Ilegales
12	3	3.0000	Policía
13	10	2.0000	Bloque Libertadores Sur
14	11	1.0000	Urabeños
15	13	1.0000	Fuerzas Armadas
16	12	1.0000	M19
17	17	1.0000	Eln
18	14	1.0000	M21

La segunda métrica estructural por analizar corresponde a Betweenness centrality. Por ello, se hace uso de la ejemplificación encontrada en la tabla **1-22**, donde se muestran los términos considerados como nodos centrales de acuerdo con un alto valor dentro de su repetición. Por lo tanto, los nodos centrales vendrían siendo las FARC, los Rastrojos, las AUC y las Autodefensas Campesinas Nueva Generación. Es interesante percibir como estas cuatro entidades se convierten en nodos centrales partiendo de valores diferentes, mientras las demás entidades obtienen puntajes más alejados.

Tabla 1-22: Betweenness centrality: Organización - Organización

Rank	Vertex	Value	Id
1	2	0.1988	Farc
2	6	0.1841	Rastrojos
3	15	0.1735	Auc
4	7	0.1173	Autodefensas Campesinas Nueva Generación
5	4	0.0357	Ejército
6	9	0.0267	Bacrim
7	5	0.0119	Águilas Negras
8	1	0.0054	Bloque Central Bolívar
9	10	0.0029	Bloque Libertadores Sur
10	8	0.0009	Grupos Armados Ilegales
11	18	0.0000	Fuerza Pública
12	13	0.0000	Fuerzas Armadas
13	17	0.0000	Eln
14	3	0.0000	Policía
15	12	0.0000	M19
16	11	0.0000	Urabeños
17	16	0.0000	Águilas Negras
18	14	0.0000	M21

La tercera métrica estructural por analizar corresponde a All closeness centrality. Esta es posible de evaluar gracias a la información recogida en la tabla **1-23**, dónde se muestran las organizaciones más cercanas y lejanas dentro del corpus. En esta ocasión, las entidades Rastrojos, Autodefensas Campesinas Nueva Generación, FARC, AUC y Ejército poseen valores muy parecidos que son los mayores, ubicándolas como las más cercanas. Para esta ocasión, es interesante percibir como de las cinco más cercanas, una hace parte de la rama ejecutiva de Colombia, poniendo en entredicho el rol de esta entidad dentro de la problemática de los actores victimizantes. Otra posibilidad de su presencia puede recaer en la necesidad de contrarrestar la fuerza ilegal que adquiere fuerza en este componente geográfico.

Tabla 1-23: All closeness centrality: Organización - Organización

Rank	Vertex	Value	Id
1	6	0.6667	Rastrojos
2	7	0.6349	Autodefensas Campesinas Nueva Generación
3	2	0.6349	Farc
4	15	0.6061	Auc
5	4	0.6061	Ejército
6	1	0.5333	Bloque Central Bolívar
7	5	0.5128	Águilas Negras
8	9	0.4938	Bacrim
9	18	0.4762	Fuerza Pública
10	8	0.4301	Grupos Armados Ilegales
11	16	0.4167	Águilas Negras
12	3	0.4040	Policía
13	10	0.4040	Bloque Libertadores Sur
14	11	0.3922	Urabeños
15	12	0.3810	M19
16	17	0.3704	Eln
17	13	0.1111	Fuerzas Armadas
18	14	0.1111	M21

Por último, la cuarta métrica estructural se compone de dos: Hubs y Authorities. Para el primer caso, se encuentra disponible la ejemplificación resaltada a través de la tabla 1-24, donde se puede rescatar a las entidades con mayor circulación dentro del corpus. Para este caso vendrían siendo la Policía y el Bloque Central Bolívar. Es bastante interesante analizar este resultado, puesto que el objetivo del informe se consagra en la línea del análisis de los actores victimizantes sobre las víctimas. Por ello, la explicación de la aparición recurrente de la policía podría deberse a su vínculo con el conflicto armado, así como su defensa de los aspectos violentos. Por otro lado, el segundo caso es representado a través de la tabla 1-25, donde se puede apreciar a las entidades con mayor circulación dentro de la red. Dentro de este caso, dichas entidades serían las Autodefensas Campesinas Nueva Generación y los Rastrojos. De esta manera, los resultados son congruentes con las pasadas métricas estructurales.

Tabla 1-24: Hubs: Organización - Organización

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	9	47.3684	9	44.4444	Policía
2	10	52.6316	19	100.0000	Bloque Central Bolívar

Tabla 1-25: Authorities: Organiación - Organización

Rank	Vertex	Value	Id
1	7	0.5271	Autodefensas Campesinas Nueva Generación
2	6	0.5149	Rastrojos
3	2	0.4307	Farc
4	17	0.2416	Auc
5	1	0.2288	Bloque Central Bolívar
6	19	0.1984	Fuerza Pública
7	18	0.1858	Águilas Negras
8	4	0.1762	Ejército
9	8	0.1330	Grupos Armados Ilegales
10	5	0.1375	Águilas Negras

De acuerdo con los diferentes resultados expuestos anteriormente de la pareja de nodos organización - organización, se pueden ejemplificar análisis más profundos. Según el informe **Núcleo Desplazamiento forzado y despojo de tierras**, “con la vida desplazada y la tierra usurpada”, en el sur occidente del país el desplazamiento es el hecho victimizante más común de la región, ya que cerca de la mitad de las víctimas fueron desplazadas. Para empezar, este hecho victimizante tiene raíces en los enfrentamientos entre grupos armados por el control del territorio, de la población y por los cultivos de coca y su eventual beneficio económico.

En el sur occidente del país, “los paramilitares se acentuaron a mediados de los años 1999 y 2000” [7]. Especialmente en el departamento de Nariño se creó y se fortaleció el grupo llamado Bloque Libertadores del Sur. Este, a su vez, tuvo una división significativa en tres bandas que se nombran a continuación: Las Autodefensas Nueva Generación, Las Águilas Negras y Los Rastrojos.

A partir de esta información se encuentra que, en efecto, estas tres bandas criminales tienen un alto valor de frecuencia en las sentencias de Nariño que fueron analizadas. Al observar detenidamente la tabla **1-21**, se encuentra que Los Rastrojos ocupan el lugar número dos, Las Águilas Negras el puesto número siete, y Las Autodefensas Nueva Generación el primer puesto. Así mismo, de acuerdo con la Fiscalía, “Las Farc tienen el mayor número de víctimas atribuidas al desplazamiento forzado en esta región” [7]. Esta organización ocupa el primer lugar en la tabla anteriormente mencionada. Estos resultados también indican el grado de influencia y posicionamiento de estas organizaciones dentro de la región; el nombre de estas se repite constantemente dentro de las sentencias, lo que quiere decir que sus acciones y repercusiones dentro de la sociedad y el territorio son importantes. El hecho de que estas organizaciones tengan un alto valor de frecuencia también señala que sus actuaciones se

produjeron más de una vez, que fueron constantes y sobre todo que no siguieron un curso natural, sino que irrumpieron con el orden establecido dentro del territorio.

Ahora bien, cuando analizamos la tabla **1-23**, aparecen de nuevo organizaciones como las Autodefensas Campesinas Nueva Generación y Los Rastrojos como las organizaciones más centrales dentro de las sentencias, y en contraste, el M21 o las fuerzas armadas como menos centrales. Lo anterior se relaciona con el informe **Núcleo Desplazamiento forzado y despojo de tierras**, ya que dentro de este se mencionan a estos grupos como los principales actores armados dentro del departamento del Nariño. Desde esta perspectiva, la tabla **1-22**, confirma estas conclusiones. A partir de la imagen **1-5**, los cuatro nodos más importantes son Auc, Farc, Rastrojos, y Autodefensas Campesinas Nueva Generación.

Así, estas organizaciones se vuelven importantes, fundamentales y decisivas a la hora de analizar las causas de desplazamiento en el departamento de Nariño. Al determinar qué actores armados se acentuaron en determinado territorio, se dilucidan los caminos para llegar a la verdad, para encontrar a los responsables, entender su modus operandi, y así mismo proponer soluciones para la no repetición; entendiendo desde el principio el quién, el cómo y el por qué.

Como conclusiones generales, es importante resaltar la gran presencia de las organizaciones ilegales, en especial, aquellas con nexos con el narcotráfico y los paramilitares. A su vez, es interesante analizar la presencia de organizaciones institucionales pertenecientes al ejecutivo, puesto que el informe se desarrolló con la intención de demostrar a las entidades victimizantes sobre las víctimas. De la misma manera, es interesante analizar los verbos presentes y su contenido violento. Esto se debe a que, si se encuentran examinando hechos enmarcados en el conflicto armado con la intención de desvelar los actores, es bastante curioso que los verbos usados en la recolección de las entrevistas presenten un bajo nivel de violencia. Verbos como adquirir, identificar, pretender o informar no dicen mucho respecto al contenido violento de la cronología y geografía analizada. Esta última se encuentra congruente con los resultados obtenidos, puesto que se encuentra analizando específicamente, de manera regional, el departamento de Nariño. Frecuencias de lugares como Municipio Policarpa y Bella Vista ayudan a sustentar esta teoría.

Por último, es importante recalcar, al igual que lo dicho en las anteriores secciones de este capítulo, que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas.

1.6. Análisis de redes de entidades presentes en sentencias de la Unidad de Restitución de Tierras (URT) Cesar

Este informe fue culminado el 25 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Este documento es el resultado del trabajo analítico del corpus de sentencias de la Unidad de Restitución de Tierras (URT). Si bien las víctimas juegan un papel demasiado importante en este informe, se busca ahondar en las personas u organizaciones que aparecen como victimarios, ya que desde este punto de vista se puede entender el cómo y el porqué del conflicto armado desde otra perspectiva. Como se puede apreciar, los objetivos principales de este informe son exactamente semejantes a los del acontecido en la sección previa, no obstante, el cambio se desarrolla en la delimitación geográfica. De esta manera, se puede analizar la diferencia de la presencia de victimarios de una región a otra. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Organizaciones - Organizaciones

- Organizaciones - Verbos

- Personas - Organizaciones

- Personas - Personas

- Personas - Verbos

Un ejemplo de la presentación de los resultados de los exámenes de estas parejas de nodos se puede encontrar en la imagen **1-6**, dónde se permite evidenciar el análisis gráfico de la pareja de nodos organizaciones - organizaciones. De acuerdo con los resultados, se pueden identificar agrupaciones no solamente ilegales sino también entidades de corte comercial y económico como es el caso de Ogx Petróleo. Además, es interesante analizar como estas relaciones no son escasas, sino todo lo contrario. Casi que todas las organizaciones de corte guerrillero o paramilitar tienen una conexión con una empresa de este tipo.

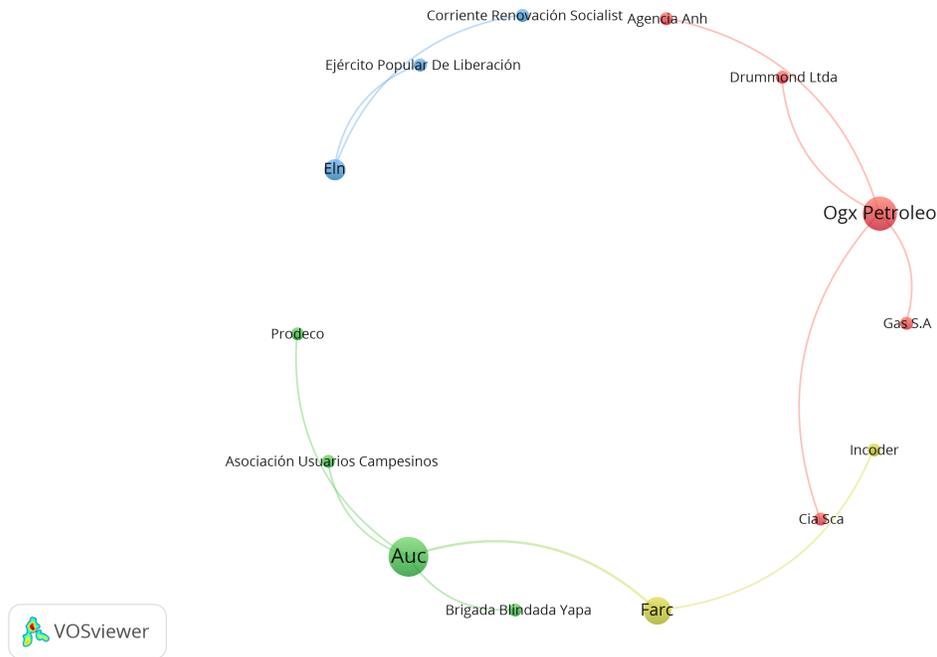


Figura 1-6: Organización - Organización

Un ejemplo de la métrica estructural All Degree puede ser apreciada a través de la tabla **1-26**, donde se evidencian los términos más frecuentes dentro del corpus. Entre estos se encuentran las AUC, Ogx Petróleo y las FARC. Es interesante analizar como dos organizaciones victimizantes famosas alcanzan a obtener una frecuencia igual que una corporativa económica.

Tabla 1-26: All Degree: Organizaciones - Organizaciones

Rank	Vertex	Value	Id
1	2	5.0000	Auc
2	10	4.0000	Ogx Petroleo
3	1	3.0000	Farc
4	5	2.0000	Eln
5	11	1.0000	Agencia Anh
6	14	1.0000	Gas S.A
7	9	1.0000	Brigada Blindada Yapa
8	8	1.0000	Prodeco
9	7	1.0000	Asociación Usuarios Campesinos
10	6	1.0000	Corriente Renovación Socialista
11	13	1.0000	Cia Sca
12	4	1.0000	Ejército Popular De Liberación
13	3	1.0000	Incoder
14	12	1.0000	Drummond Ltda

Ahora bien, para el caso de la métrica estructural del Betweenness centrality, es importante destacar como en la tabla **1-27** se obtienen los siguientes nodos centrales: AUC, Ogx Petróleo y FARC. Como se puede apreciar, las mismas organizaciones consideradas más frecuentes dentro del corpus, a su vez, componen a los nodos centrales.

Tabla 1-27: Betweenness centrality: Organizaciones - Organizaciones

Rank	Vertex	Value	Id
1	2	0.1154	Auc
2	10	0.0769	Ogx Petroleo
3	1	0.0513	Farc
4	5	0.0128	Eln
5	11	0.0000	Agencia Anh
6	14	0.0000	Gas S.A
7	9	0.0000	Brigada Blindada Yapa
8	8	0.0000	Prodeco
9	7	0.0000	Asociación Usuarios Campesinos
10	6	0.0000	Corriente Renovación Socialista
11	13	0.0000	Cia Sca
12	4	0.0000	Ejército Popular De Liberación
13	3	0.0000	Incoder
14	12	0.0000	Drummond Ltda

Un ejemplo de la métrica estructural All closeness centrality se encuentra representada en la tabla **1-28**, dónde se pueden apreciar las siguientes entidades: AUC, Ogx Petróleo y FARC. Estas organizaciones representan valores de cercanía bastante altos. A partir del cuarto término, los valores de cercanía empiezan a parecer bastante cercanos unos entre otros. De todas maneras, la congruencia con las previas secciones se mantiene.

Tabla 1-28: All closeness centrality: Organizaciones - Organizaciones

Rank	Vertex	Value	Id
1	2	0.3571	Auc
2	10	0.3571	Ogx Petroleo
3	1	0.2679	Farc
4	7	0.2143	Asociación Usuarios Campesinos
5	5	0.2143	Eln
6	9	0.2143	Brigada Blindada Yapa
7	8	0.2143	Prodeco
8	12	0.2041	Drummond Ltda
9	14	0.2041	Gas S.A
10	13	0.2041	Cia Sca
11	11	0.2041	Agencia Anh
12	3	0.1786	Incoder
13	6	0.1429	Corriente Renovación Socialista
14	4	0.1429	Ejército Popular De Liberación

Por último, se encuentran las métricas estructurales de Hubs y Authorities. Para el primer caso se encuentra la tabla **1-29**, la cual representa las entidades que más circulan dentro del corpus, tales como Ejército Popular de Liberación y las FARC. En este caso se evidencia que se rompe la continuidad respecto con las anteriores tres métricas estructurales, dejando a las FARC como la única organización que se mantiene presente. Por otra parte, para el segundo caso, se encuentra la tabla **1-30**, la cual representa las entidades que más circulan dentro de la red, tales como las AUC, FARC e Incoder. En esta ocasión, los resultados son congruentes con lo visto en las previas métricas estructurales, no obstante, a la inclusión de Incoder resulta curioso al ser el Instituto Colombiano de Desarrollo Rural.

Tabla 1-29: Hubs: Organizaciones - Organizaciones

Cluster	Freq	Freq %	CumFreq	CumFreq %	Representative
0	4	28.5714	4	28.5714	Ejército Popular de Liberación
2	10	71.4286	14	100.0000	Farc

Tabla 1-30: Authorities: Organizaciones - Organizaciones

Rank	Vertex	Value	Id
1	2	0.7924	Auc
2	1	0.4483	Farc
3	3	0.2399	Incoder
4	7	0.1947	Asociación Usuarios Campesinos
5	9	0.1947	Brigada Blindada Yapa
6	8	0.1947	Prodeco
7	4	0.0000	Ejército Popular De Liberación
8	14	0.0000	Gas S.A
9	13	0.0000	Cia Sca
10	6	0.0000	Corriente Renovación Socialista

Por lo tanto, en estos ejemplos se presentaron las organizaciones más y menos importantes dentro del corpus. En primera instancia aparecen las Farc, las Auc, el Ejército, el Eln, El Incoder, Prodeco y la Brigada Blindada Yapa. Llama la atención el nombre de Prodeco que hace referencia a “la empresa estatal de explotación del carbón mineral que fue adquirida por C.I. Prodeco S.A., filial de la suiza Glencore” [7].

Si bien esta empresa no está vinculada directamente con procesos de desplazamiento, sí parece estar vinculada con grupos paramilitares como el frente Juan Andrés Álvarez. Sumado a lo anterior, esta empresa compró terrenos a terceros que obligaron a propietarios campesinos a vender sus tierras en un precio que no correspondía al valor real. Lo anterior se sustenta en confesiones de Salvatore Mancuso, uno de los jefes principales de este frente.

En la tabla **1-27**, la cual muestra los nodos más importantes, aparece el Incoder. Esta entidad estatal está vinculada con problemas en la titulación de tierras a campesinos en el Cesar. Especialmente en la parcelación El Prado, “el Incoder y un notario de la zona despojaron a las personas que vivían allí” [7]. Para lograr su cometido, el Incoder falsificó “títulos de propiedad y su entrega a terceros que resultaron ser familiares y testaferros de los paramilitares” [7]. De esta manera, el Incoder traspasó varias parcelas a familiares del ex paramilitar David Hernández. No siendo esto suficiente, para el 2011 mediante un fallo del Juzgado Penal del Circuito Especializado de Valledupar se le ordenó a la multinacional Prodeco S.A y al Incoder “restituir los derechos de posesión de las tierras y convertir en propietarios a los parceleros de El Prado” [7]. Sin embargo, hoy en día las víctimas siguen esperando que se les restituyan sus territorios de los que fueron desplazados violentamente hace diez años.

Por otro lado, teniendo en cuenta que el Cesar es un territorio minero, fue debido a la ex-

pansión de proyectos de este tipo que muchos campesinos fueron violentamente despojados de sus tierras. Es el caso de los parceleros del Toco quienes “fueron víctimas de masacres y asesinatos cometidos por paramilitares de las AUC” [7]. Como resultado, se llevó a cabo un “proceso de despojo territorial en cabeza de beneficiarios del paramilitarismo, afianzando la minería de carbón en este territorio” [7]. Esta última organización ocupa un lugar importante dentro de las sentencias de Cesar, pues es la más importante según la tabla **1-26**, el primero más central de acuerdo con la tabla **1-27** y la más central en el corpus siguiendo a la tabla **1-28**.

Por último, al pasar a la tabla **1-29**, la cual señala a las organizaciones que tiene mayor circulación en el corpus, se muestra que las Farc ocupan el segundo lugar. Este grupo armado ha hecho presencia en la región del Cesar desde los años ochenta con los frentes “Camilo Torres Restrepo, el frente 6 de diciembre y el frente 19” [7]. Los actores delictivos de esta organización eran varios: “extorsionaban y secuestraban a los dueños de las haciendas ganaderas y agricultoras y realizaban atentados con bombas a algunas fincas que se negaban a sus pretensiones” [7].

Así pues, las organizaciones que aparecen en este apartado han contribuido a la violencia, desplazamiento y guerra en el Cesar. Estas organizaciones han encontrado en esta región una gran oportunidad económica debido a los territorios y a la forma en que estos son explotados. Cabe resaltar el vínculo entre unas y otras y los hechos victimizantes de los que han sido autores en el Cesar.

Como conclusiones generales, es importante establecer diferencias claras entre este informe y el anterior, puesto que, aunque tratan temáticas iguales con los mismos enfoques de victimizantes, estos se desarrollan en dos espacios geográficos diferentes. Es por ello que, por ejemplo, los mayores actores victimizantes de esta región se distribuyan entre guerrilleros, paramilitares, entidades gubernamentales y corporaciones. Del mismo modo, es interesante ver que, en este caso, se presenta una similitud con el informe anterior. Esto se debe a que la presencia de los verbos analizados no hace parte de hechos victimizantes. Por el contrario, se desarrollan acciones de adhesión a una organización o acciones transitivas con ayuda de verbos como tener o hacer.

Por último, es importante recalcar, al igual que lo dicho en las anteriores secciones de este capítulo, que gran parte de las frecuencias de algunas de las categorías corresponden a nombres propios carentes de contexto de análisis y a expresiones dialécticas recogidas dentro del discurso narrativo de las entrevistas.

Bibliografía

- [1] A.-L. Barabási, “Network science,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.
- [2] P. González, Andrea & Gómez Valentina & Amado, “Masacres,” 2021.
- [3] R. A. Española, “Diccionario de la real academia española,” *Diccionario de la lengua española*, vol. Diccionario Online, no. 1, s.f.
- [4] Fiscalía General de la Nación, “Ley de justicia y paz, compilación normativa y jurisprudencia,” Jul. 25 2005. https://www.fiscalia.gov.co/colombia/wp-content/uploads/2012/04/ley_975_de_2005.pdf.
- [5] Verdadabierta, “El comandante “narco” del frente barro blanco de tarazá,” Jul. 23 2013 [Online]. <https://verdadabierta.com/el-comandante-narco-del-frente-barro-blanco-de-taraza/>.
- [6] E. C. Rodríguez and S. T. Silva, “El movimiento político y social marcha patriótica: génesis, estructura y proyecto,” *Revista Colombiana de Ciencias Sociales*, vol. 8, no. 1, pp. 189–225, 2017.
- [7] C. de la verdad, ““con la vida desplazada y la tierra usurpada” informe nucleo desplazamiento forzado y despojo de tierras,” vol. Informe sin publicar, no. pp. 1 - 46, 2021.

2 Relaciones y Coocurrencias

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las entrevistas de los diferentes corpus utilizados dentro de los diversos informes que comprenden múltiples objetivos y preguntas de investigación. Esto con el fin de establecer relaciones y coocurrencias lingüísticas. Este análisis se describe con los resultados automáticos del componente de software que el grupo de Procesamiento Lenguaje Natural desarrolló en una API, el cual tiene el nombre “Noun Phrases” para un análisis de entramado realizando tres tipos de gráficas y grafos sobre un grupo de interés del corpus ingresado.

Los corpus procesados contienen un diferente número de entrevistas, no obstante, la metodología continúa siendo la misma. Cuando se tokenizan las entrevistas de la plataforma para su análisis, se encuentran dentro de ellas distintos términos irrelevantes denominados “stop-words”, los cuales se limpian de los resultados parciales, así como los caracteres especiales como los paréntesis, tildes, signos de puntuación, etc. Una vez son filtrados los resultados, se procesan a través de una herramienta de conteo utilizando las librerías Spacy y Matplotlib, que se explicarán a continuación, arrojando una categorización de cada término del corpus según los miembros de las parejas de nodos de interés.

Spacy es una librería de software para procesamiento de lenguajes naturales desarrollado por Matt Honnibal y programado en lenguaje Python. Su utilidad en la propuesta es la capacidad que otorga para tokenizar, es decir, dividir el texto en unidades lingüísticas individuales de manera automática [1]. De esta manera, Spacy nos permite reconocer la categoría gramatical de una palabra dada (verbo, adjetivo, etc.), además de filtrar su lema, lo que hace más fácil la identificación sin importar la flexión verbal en la que se encuentre.

Matplotlib es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python. Fue utilizada para proporcionar resultados gráficos de los datos analizados. Estos gráficos son de barras, matrices de ocurrencias y nubes de palabras.

El primer tipo de gráfico combina un diagrama de barras y la nube de palabra para realizar un estudio de las frecuencias de palabras sobre un elemento del grupo de interés. El diagrama de barras contiene el top 10 de las palabras que fueron catalogadas dentro de ese grupo y la nube de palabras representa la mayor frecuencia de la palabra en un mayor tamaño de dicha

palabra.

El segundo tipo de gráfico es una matriz de coocurrencias entre una combinación (par) de los miembros del grupo de interés; por ejemplo, las relaciones que pueden tener los verbos y los adjetivos en aras de ver las calificaciones de los actos cometidos durante el conflicto armado. En síntesis, el gráfico está orientado en un mapa de calor con el objetivo de identificar y calcular las relaciones más relevantes; es decir, cuanto mayor sea el tono del color, se interpreta que hay una mayor relación entre una categoría y la otra.

Finalmente, el trabajo metodológico comprende grafos estructurales de redes para visualizar las relaciones entre los grupos de interés; en este caso, los grafos resultantes de la API fueron modificados con un software complementario VosViewer. Así pues, se realizó una limpieza de los resultados parciales para la construcción de redes que no sean genéricas sino más específicas con términos que apunten a relaciones relevantes y/o interesantes. Los resultados de las redes son evaluados a partir de la siguiente métrica estructural que se explica sucintamente así:

All degree: Señala los términos más y menos importantes en el corpus con su valor de frecuencia, es decir, representa un número de enlaces que se tiene con otros nodos [2]. Su medida se calcula con la siguiente expresión:

$$C_D(i) = \frac{1}{2} \sum_{i=1}^N k_i$$

Dónde k_i representa el nodo existente dentro de la red.

A continuación, se presentarán secciones respecto a los diferentes informes que utilizan esta metodología dentro de su análisis.

2.1. Análisis de entramado sobre las entrevistas colectivas

Este informe fue culminado el 15 de marzo de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. En el presente informe se detalla el trabajo de procesamiento de corpus de las entrevistas recolectadas por la Comisión de la Verdad etiquetadas con el tema de Colectivas. Dentro de las mismas, se busca encontrar patrones y conexiones representadas por sistemas de redes que permitan generar un estudio a mayor profundidad dentro de Colectivas. Dentro de los hallazgos presentados en las redes, es importante recalcar que la delimitación cronológica del trabajo se encuentra entre 1950 y 2021. Del mismo modo, la delimitación geográfica es bastante amplia, ya que se tienen en cuenta departamentos, ciudades, municipios y corregimientos en los cuales se datan las entrevistas. Esta delimitación

geográfica es lo suficientemente extensa para cubrir todos los departamentos del país.

A través del procesamiento del corpus, se encontró un resultado promedio de 298.9 tokens aproximadamente con una diversidad lexical promedio de 0.5 aproximadamente. De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Chunks - Adjetivos
- Chunks - Organizaciones
- Chunks - Personas
- Chunks - Lugares
- Chunks - Verbos
- Organizaciones - Organizaciones
- Organizaciones - Lugares
- Organizaciones - Verbos
- Personas - Organizaciones
- Personas - Personas
- Personas - Verbos

A continuación, se presentan los diferentes gráficos expuestos en la metodología del capítulo, poniendo como ejemplo la frecuencia de sustantivos – adjetivos y, para las otras dos gráficas, la pareja de nodos Organización – Organización.

La imagen **2-1** permite identificar las frecuencias de los diez términos más relevantes dentro del corpus que componen la categoría de sustantivos – adjetivos. Entre estos se encuentran conflicto armado, grupo armado, actor armado, fuerza pública entre otros. Por lo tanto, es posible inferir la gran presencia de adjetivos como armado dentro del corpus al ser una palabra clave en el desarrollo de la construcción lingüística de las narrativas dentro de las entrevistas. Del mismo modo, se permite evidenciar la característica influyente en la cual se encuentra enmarcado el trabajo: el conflicto armado, quienes lo prolongan y desde donde. Al mismo tiempo, la segunda parte de la imagen, la cual se encuentra en la parte derecha, evidencia la misma información que las barras de frecuencia, con la única diferencia de que es más inclusivo. Esto significa que se encuentran más términos además de los diez más frecuentes. La forma de presentación es con las nubes de palabras que permiten identificar la relevancia de un término entre mayor sea su tamaño.

escenario para participar y exigir sus derechos, puesto que el Estado colombiano reconoció que los daños producidos por el conflicto armado trascendieron las afectaciones individuales e impactaron el conjunto de la organización” [6].

Ahora bien, para el caso de los grafos construidos a partir de VosViewer con la pareja de nodos de organización – organización, se produce la imagen **2-2**, la cual presenta las organizaciones más frecuentes o relevantes dentro del corpus. En esta ocasión, se presentan organizaciones guerrilleras, paramilitares, institucionales y políticas, dejando entrever que las relaciones dentro del corpus se componen desde diferentes sectores. No obstante, es importante aclarar que estas relaciones no son simétricas, puesto que hay algunas entidades con mayor frecuencia y un mayor número de relaciones, como es el caso de las FARC, el EPL o la Comisión de la Verdad. Como se evidencia, las relaciones apuntan más a un desarrollo con grupos ilegales con algunos interventores como la comisión en la recolección de testimonios. No obstante, la sustancia de la investigación se encuentra en las acciones violentas y dolientes cometidas desde un sector militante que se mantienen en las psiquis de los interlocutores.

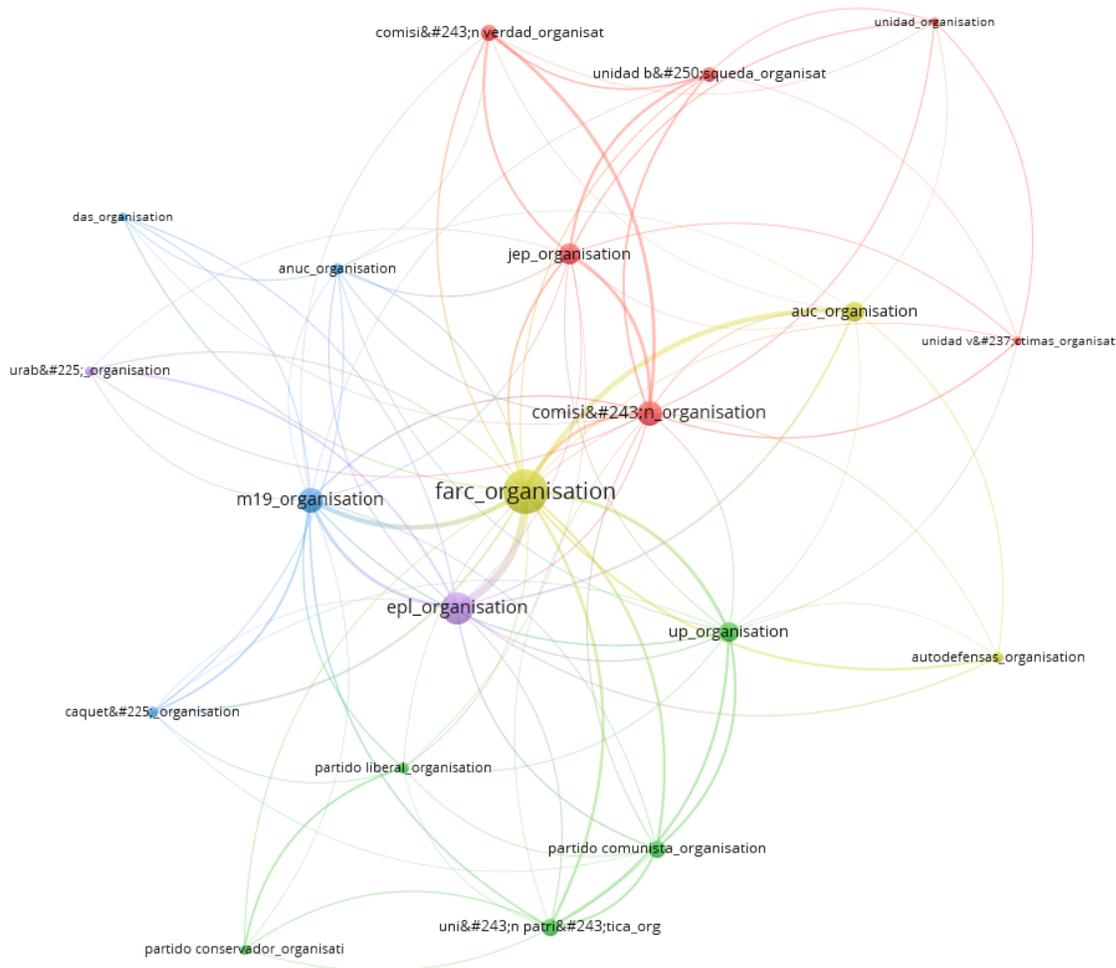


Figura 2-2: Network lecture Organizaciones - Organizaciones

Una relación más profunda relacionada entre las FARC y el EPL se puede entrever por los enfrentamientos entre estos dos grupos en dos instancias: en la década de los 80 y en la década de los 90, donde el EPL había firmado los acuerdos de paz y sólo una fracción disidente era la que combatía con las FARC. Los combates se desarrollaron en la región del Urabá a causa de una disputa territorial y por el control del sindicato bananero, además de la desertión de los núcleos marxistas-leninistas de las FARC y su adhesión al EPL [7, p. 38]. Por último, la explicación de manera más profunda de la relación entre las FARC y el M19 proviene de ámbitos menos combativos, puesto que la relación de estas guerrillas, en la década de los setenta, se presentaba más como una llamada a la unión entre organizaciones, tarea impulsada en la década de los ochenta por parte de las FARC y el M19 [7, p. 39-40]. Del mismo modo, las FARC no atacaron a los desmovilizados guerrilleros provenientes de la guerrilla del M19. Únicamente se puede rescatar un enfrentamiento entre ambas guerrillas, no obstante, dicho enfrentamiento se daría en función de la disputa política por acercamientos

de paz diferenciados por el gobierno entre 1980 y 1986 [7, p. 38].

Para terminar, para el caso de las gráficas desarrolladas a partir de las coocurrencias entre los nodos organización - organización, se puede encontrar la imagen **2-3**. En esta se presentan coocurrencias con valores de hasta 5, no obstante, estos son escasos. De hecho, estos casos corresponden a coocurrencias entre las mismas organizaciones, es decir, por ejemplo, las FARC con las FARC. De igual manera, un número elevado de coocurrencias se desarrollan entre entidades y errores dialécticos dentro del corpus con expresiones como “Pp” o “acá”. No obstante, aun se presentan algunas coocurrencias con valores elevados, como es el caso de las FARC con las organizaciones AUC, EPL y M-19.

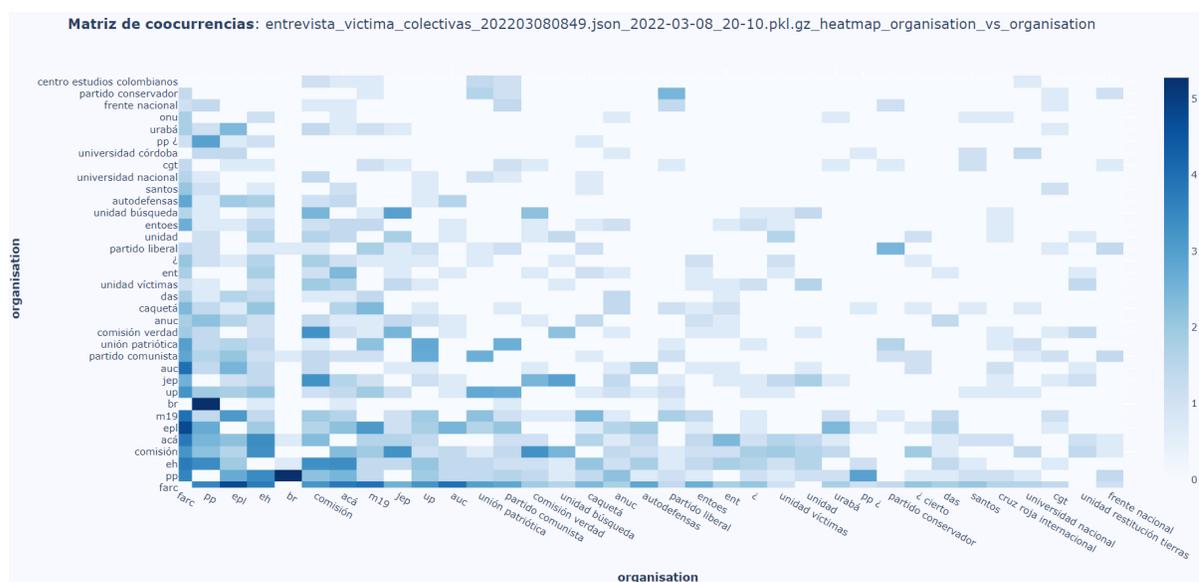


Figura 2-3: Diagrama Organizaciones - Organizaciones

Una relación más profunda entre las FARC y las AUC, a parte de los paralelos realizados referentes a los procesos de desmovilización de estos dos grupos, proviene de los enfrentamientos entre estas dos organizaciones en los departamentos de Cauca y Nariño. Una confrontación que se desenvolvería durante la década de los noventa y la primera década del siglo XXI, cuando se desarrolló una gran desmovilización de varias estructuras paramilitares de las zonas. Esto provoca que las FARC ocupen los territorios previamente de los paramilitares y ajusticiaran a las personas sospechosas de haber ayudado a las autodefensas [7, p. 80]

Como conclusiones generales, las mayores relaciones y coocurrencias encontradas dentro del corpus se componen de un denominador común: las Fuerzas Armadas Revolucionarias de Colombia (FARC). Esto se debe a que se encuentran conectadas con los sintagmas nominales, las personas, las organizaciones y los verbos más frecuentes dentro del corpus. Del mismo modo, estas relaciones que se desencadenan de las FARC son las más fuertes dentro de los análisis de los diagramas y los network lectures, algún otro término no llega a presentarse con la misma frecuencia dentro de las relaciones y coocurrencias. Aquellos más

cercanos a las FARC o que poseen una frecuencia y fuerza de relaciones igual o mayor son en realidad expresiones discursivas dentro de la narrativización de las entrevistas. Expresiones como “pp”, “etoes”, “3 año”, entre otros. Por lo tanto, es importante la filtración de este tipo de datos al realizar las frecuencias, diagramas y los network lectures, puesto que afectan de manera negativa el análisis de los resultados. Del mismo modo, es importante concluir respecto con los datos encontrados de manera incorrecta referenciada en las categorías. Estos datos son aquellos que poseen, por ejemplo, la categoría de organización, cuando en realidad son verbos. Este tipo de datos lo que generan es una inflación de los resultados de las frecuencias dentro del corpus y, por ende, presentan con relaciones más fuertes caracteres de los corpus analizados cuyas relaciones, en realidad, no sean tan fuertes desde un principio.

Por último, es importante mencionar que gran parte de las mayores fuerzas en las relaciones y coocurrencias encontradas se desarrollan con los verbos. Esto se debe a que, en cada oración presentada en los discursos narrativos, es importante la generación de un vínculo con lo experimentado por el narrador o la narradora, desencadenando una confrontación con el pasado vuelto acontecimiento presente. Es decir, que los verbos trascienden sobre todas las demás categorías al ser el insumo principal de la construcción del discurso narrativo de las entrevistas.

2.2. Análisis de entramado sobre las entrevistas a sujetos colectivos

Este informe fue culminado el 19 de abril de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. En el presente informe se detalla el trabajo de procesamiento de corpus de las entrevistas recolectadas por la Comisión de la Verdad etiquetadas con el tema de Sujetos Colectivos. Dentro de las mismas, se busca encontrar patrones y conexiones representadas por sistemas de redes que permitan generar un estudio a mayor profundidad dentro de los resultados de sujetos colectivos. Dentro de los hallazgos presentados en las redes, es importante recalcar que la delimitación cronológica del trabajo se encuentra entre 1492 y 2013. La explicación de este aspecto de larga duración se debe a que se encuentran entrevistas con temas que rescatan un componente colonial importante acontecido durante la ocupación española en el territorio colombiano. Del mismo modo, la delimitación geográfica es bastante amplia, ya que se tienen en cuenta departamentos, ciudades, municipios y corregimientos en los cuales se datan las entrevistas. Esta delimitación geográfica es lo suficientemente extensa para cubrir todos los departamentos del país.

De acuerdo con la metodología aplicada, las parejas de nodos desarrolladas para analizar los datos del corpus fueron los siguientes:

- Chunks - Adjetivos

- Chunks - Organizaciones

- Chunks - Personas

- Chunks - Lugares

- Chunks - Verbos

- Organizaciones - Organizaciones

- Organizaciones - Lugares

- Organizaciones - Verbos

- Personas - Organizaciones

- Personas - Personas

- Personas - Verbos

A continuación, se presentan los diferentes gráficos expuestos en la metodología del capítulo, poniendo como ejemplo la frecuencia de organizaciones y, para las otras dos gráficas, la pareja de nodos Persona - Persona.

La imagen 2-4 permite identificar las 10 organizaciones más frecuentes dentro del corpus analizado. Es importante anotar que, para estas diez más relevantes, únicamente seis hacen parte de la categoría organización. Del mismo modo, únicamente las FARC y la Comisión obtienen valores altos, mientras que otras entidades como la JEP, AUC y el EPL poseen una gran disparidad de frecuencias en relación con las FARC. Por lo tanto, los demás términos más frecuentes son en realidad errores dialécticos o muletillas mal clasificadas como organizaciones. En la segunda parte del gráfico, a mano derecha, se puede ver la misma información descrita, pero en una nube de palabras que cubre más términos. De esta manera queda más clara la lejanía de frecuencias presentes en este corpus en relación con las organizaciones.

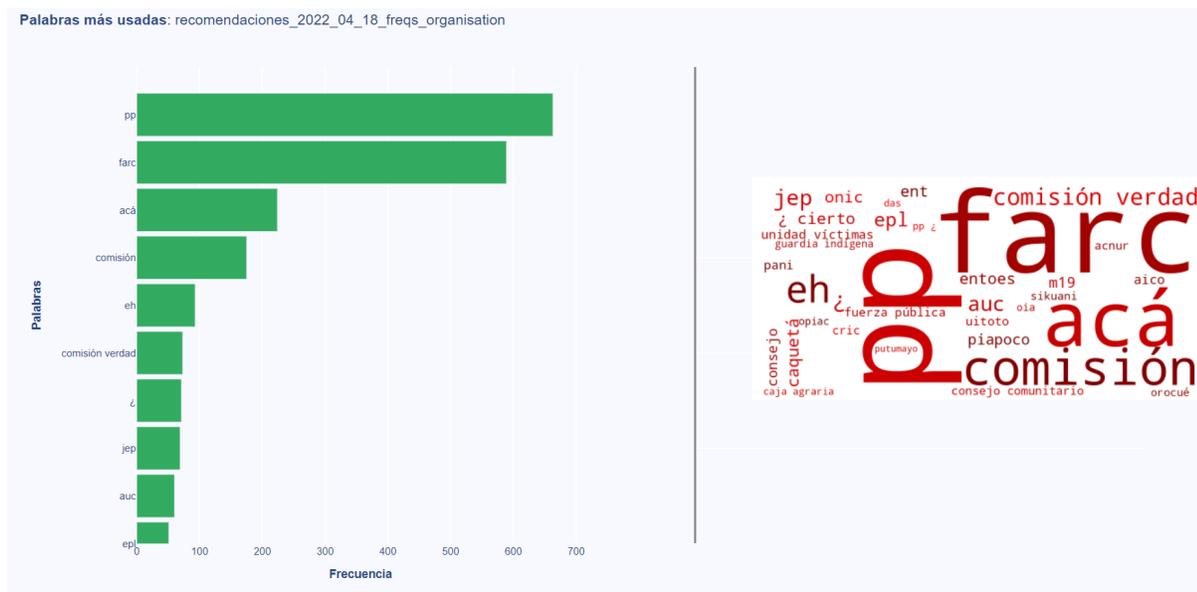


Figura 2-4: Frecuencia de organizaciones

Dentro de los análisis más profundos, es importante recalcar que únicamente se dará espacio en esta sección a los análisis de las FARC y de la Comisión de la Verdad, puesto que la frecuencia de estas organizaciones es bastante mayor que las de la JEP, AUC y EPL. La organización de las FARC se repite un total de 589 veces, mientras que la comisión (juntando comisión verdad y comisión) se repite 248 veces. El resto de organizaciones se repiten menos de cien veces.

Las FARC es algo más profundo que una organización guerrillera, se podría identificar como “una organización híbrida porque responde a motivaciones políticas y económicas, que tiene por tanto funciones políticas, criminales y militares [en otras palabras] insurgencias comerciales” [8]. Por lo tanto, al hablar de la organización de las FARC se deben de tener en cuenta las cualidades presentadas anteriormente para no caer en el argumento reduccionista de una consolidación guerrillera surgida en los años sesenta.

La Comisión de la Verdad es una organización que: “en el marco del Acuerdo Final para la terminación del conflicto y la construcción de una paz estable y duradera, suscrito entre el Gobierno de Colombia y las Fuerzas Armadas Revolucionarias de Colombia – Ejército del Pueblo FARC -EP, mediante el Acto Legislativo 01 de 2017 y el Decreto 588 de 2017, se creó la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición, como un mecanismo de carácter temporal y extrajudicial del Sistema Integral de Verdad, Justicia, Reparación y No Repetición - SIVJRNR, para conocer la verdad de lo ocurrido en el marco del conflicto armado y contribuir al esclarecimiento de las violaciones e infracciones cometidas durante el mismo y ofrecer una explicación amplia de su complejidad a toda la sociedad”

los nodos personas - personas, se puede encontrar la imagen 2-6. En esta se presentan coocurrencias con valores de hasta 2.5, no obstante, estos son escasos. Los términos con las coocurrencias más altas terminan por relacionar ítems que en realidad no hacen parte de la categoría o el nodo analizado. Tal es el caso de la coocurrencia entre “¿” y “spanp”, los cuales no son personas y han sido mal clasificados en esta sección. No obstante, la situación es más grave, puesto que las coocurrencias del nodo personas – personas analiza más que nada cualquier otro tipo de categorías que se encuentran en el corpus. De los 35 términos analizados, únicamente 3 corresponden a personas, además de que sus coocurrencias entre si son bastante precarias.

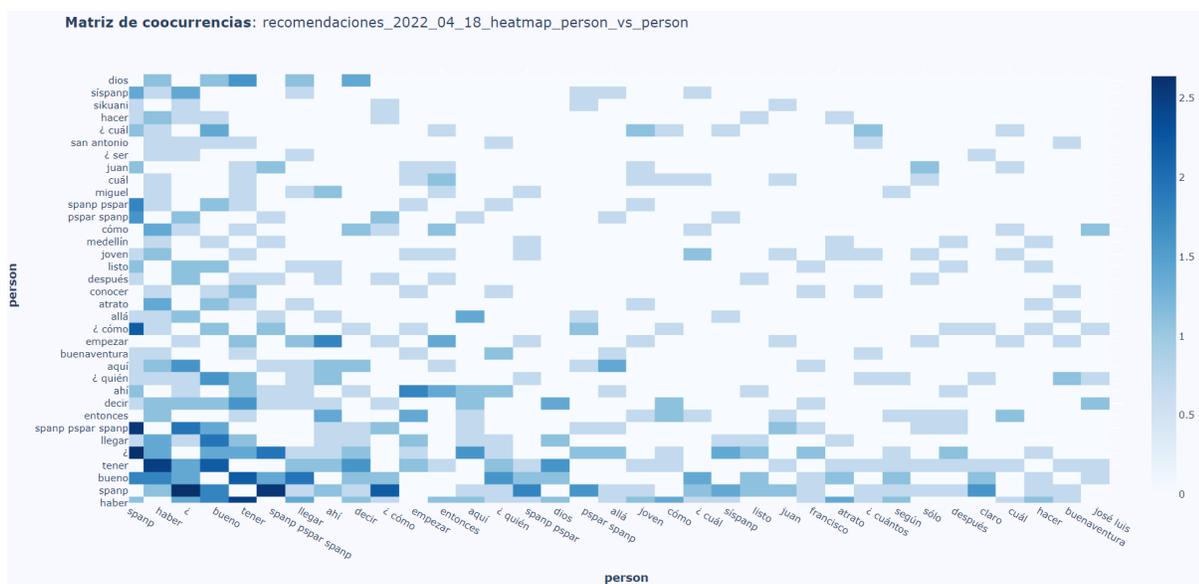


Figura 2-6: Diagrama personas - personas

Dentro de las relaciones más interesantes de la persona “Pedro” se encuentran las personas “Élmer Cárdenas” y “Pablo Escobar”. La relación con la primera persona no es del todo clara, puesto que el nombre no se desarrolla en un contexto amplio que permita un análisis adecuado. Por lo tanto, no se puede saber con seguridad si se trata de una persona por nombre Pedro o por su alias. No obstante, realizando algunas conjeturas, podría estar hablando de Pedro Parra, miembro de la conformación de las autodefensas contra las FARC en la década de los ochenta [10, p. 106]. Del mismo modo, puede tratarse de Manuel Marulanda Vélez, máximo dirigente de las FARC cuyo nombre real era Pedro Antonio Marín. Relacionándolo con Pablo Escobar podrías hasta tratarse de Pedro Gil.

Por otra parte, la imagen 2-6 presenta las coocurrencias entre personas dentro del corpus. De acuerdo con lo que se puede presenciar en la figura, se pueden identificar tres tipos de resultados. El primero de ellos corresponde a personas mal clasificadas, es decir, que dentro del análisis del corpus se pueden encontrar personas que no son personas sino otro tipo de

categoría como verbos, adjetivos o sintagmas nominales. Tal es el caso de, por ejemplo, el verbo “haber”. El segundo tipo de resultado corresponde a errores dentro del análisis del corpus al emplear palabras que no poseen ningún tipo de sentido narrativo y que no aportan nada a un análisis de personas. Entre estos dos primeros tipos de resultados, en la mayor parte, las cocurrencias son más variables. El tercer tipo de resultado corresponde a personas etiquetadas correctamente.

En conclusión, se pueden determinar las mayores relaciones y cocurrencias por categorías, resaltando a sobremanera las categorías de sintagma nominal, organización, verbo y lugar. Para la primera categoría, el ítem Conflicto armado fue esencial, puesto que poseía una alta frecuencia dentro del corpus, ocupando siempre altas relaciones. Para la segunda categoría, imperan las FARC, posicionándose siempre como la más frecuente, seguida de cerca por la Comisión. Del mismo modo, siempre se presentó, para la tercera categoría, una alta presencia dentro del corpus de los verbos decir y tener. Por último, para la categoría de lugar, Colombia y Bogotá siempre fueron centrales en los análisis. Como se puede apreciar, entre estos ítems también se desarrollaron varias relaciones dependiendo el análisis realizado en la sección. Es importante también concluir que la circulación y relación de algunas categorías superaba a otras. Tal es el caso de los verbos, los adjetivos y los sintagmas nominales. Esto se explica por sus cocurrencias que llegaban a valores de 5, mientras que categorías como personas o lugares únicamente llegaban a valores de 3, sin mencionar los valores relevantes dentro de una limpieza dentro del corpus.

Por último, es importante mencionar que gran parte de las mayores fuerzas en las relaciones y cocurrencias encontradas se desarrollan con los verbos. Esto se debe a que, en cada oración presentada en los discursos narrativos, es importante la generación de un vínculo con lo experimentado por el narrador o la narradora, desencadenando una confrontación con el pasado vuelto acontecimiento presente. Es decir, que los verbos trascienden sobre todas las demás categorías al ser el insumo principal de la construcción del discurso narrativo de las entrevistas.

Bibliografía

- [1] Oxford, “Oxford languages,” 2021.
- [2] A.-L. Barabási, “Network science,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.
- [3] O. L. V. et al., “Vinculación a grupos armados: un resultado del conflicto armado en colombia,” *Diversitas: Perspectivas en Psicología*, vol. 6, julio-diciembre 2010 [Online]. <https://www.redalyc.org/articulo.oa?id=67915140015>.
- [4] C. T. D. et al., “Niños, niñas y adolescentes en medio del conflicto armado una sistematización de las investigaciones,” 2014 [Online]. <http://polux.unipiloto.edu.co:8080/00002145.pdf>.
- [5] J. Moreno, “Conflicto armado e identidad militar en colombia 1964 - 2010,” *Revista Análisis Internacional*, vol. 20, no. 6, 2012 [Online]. <https://expeditiorepositorio.utadeo.edu.co/bitstream/handle/20.500.12010/9874/12494.pdf?sequence=1>.
- [6] J. de Acción Comunal Colombia, “La voz y el valor de las juntas de acción comunal,” [Online] <https://www.mapp-oea.org/hechosdepaz/la-voz-y-el-valor-de-las-juntas-de-accion-comunal/#:~:text=Bajo%20el%20marco%20de%20la%20Ley%201448%20del,%C3%A Dctimas%20del%20conflicto%20y%20sujetos%20de%20reparaci%C3%B3n%20colectiva.?msclkid=7ae10852abb111ec9bd4dd0c96c0cc72>.
- [7] L. M. B. Roa, “La relación entre las farc y el eln: Guerra sin cuartel y confraternidad revolucionaria (2005-2010),” 2016 [Online]. <https://repositorio.unal.edu.co/bitstream/handle/unal/59463/1023916755.2017.pdf?sequence=1&isAllowed=y>.
- [8] Óscar Palma, “¿qué han sido las farc en realidad?,” 2016 [Online] <https://www.urosario.edu.co/UCD/Que-han-sido-las-Farc-en-realidad/Documento/pdf-fasciculo-FARC-universidad-del-rosario/>.
- [9] Comisión de la verdad, “¿qué es la comisión de la verdad?,” sf [Online] <https://comisiondelaverdad.co/la-comision/que-es-la-comision-de-la-verdad>.

-
- [10] D. A. R. Rodríguez, “Historia reciente del fenómeno paramilitar (1994-2007): las historias del bloque Élder cárdenas y central bolívar de las auc,” 2015 [Online]. <https://repositorio.unal.edu.co/bitstream/handle/unal/56955/80075392.2016.pdf?sequence=1&isAllowed=y>.

3 Dimensiones

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de los corpus de entrevistas específicos de cada informe de acuerdo con unos objetivos preestablecidos haciendo uso de diferentes mecanismos de procesos de Lenguaje Natural.

El trabajo metodológico se comprende por medio del procesamiento de entrevistas pertenecientes al corpus recolectado por la Comisión de la Verdad. El procesamiento de dicho corpus se dividió entre el preprocesamiento y el procesamiento. Al ser presentado un texto dado, en el apartado del preprocesamiento, se continúa a analizar este texto por medio de librerías procedentes del lenguaje natural como lo son Spacy y Matplotlib, que se explicarán a continuación.

Spacy es una librería de software para procesamiento de lenguajes naturales desarrollado por Matt Honnibal y programado en lenguaje Python. Su utilidad en nuestra propuesta es la capacidad que otorga para tokenizar, es decir, dividir el texto en unidades lingüísticas individuales de manera automática [1]. De esta manera, Spacy nos permite reconocer la categoría gramatical de una palabra dada, además de filtrar su lema, lo que hace más fácil la identificación sin importar la flexión verbal en la que se encuentre.

Matplotlib es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python. Fue utilizada para proporcionar resultados gráficos de los datos analizados.

Una vez definidos los parámetros y resultados del preprocesamiento, se continúa con el procesamiento de los datos suministrados como se muestra a continuación.

Get_cruelty: Esta fue la función principal escrita por el equipo de analítica, en el lenguaje de programación Python, y qué está diseñada para, una vez establecidos dentro de la máquina el corpus y el material de entrenamiento, crear una serie de listas de información relacionadas en tablas y gráficos respectivamente para posibilitar así el análisis. La función a su vez elimina las entrevistas que no arrojan un resultado coincidente con los criterios mencionados.

Graficación: A partir de los resultados proporcionados por el preprocesamiento y la función explicada anteriormente, los datos se grafican de acuerdo con tres diferentes mecanismos:

- Frecuencia en barras: la cual demuestra la cantidad de veces que aparece una categoría específica en un año específico. Esta categoría puede ser desde verbos hasta sintagmas nominales.
- Promedios índices: la cual demuestra la referencia a los grados correspondientes a la categoría analizada en la frecuencia de barras respecto a una escala de interés desarrollada para cada informe.
- Caja de Bigotes: el cual demuestra los promedios de acuerdo con los años analizados en función de las categorías y dimensiones preestablecidas, lo cual ayuda a desvelar los datos que se encuentran más alejados y no acordes con el corpus.

Por último, es importante mencionar que para cada informe se construye o se usa un lexicón que proporciona luces respecto a la metodología a aplicar, puesto que de este lexicón surgen las categorías que se analizan, junto con las escalas de interés.

3.1. Análisis automatizado de corpus: indiciabilidad y dimensión de sevicia utilizando herramientas de procesamiento del lenguaje natural

Este informe fue culminado el 26 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Como parte de los intereses de la Comisión de la Verdad, y en su deber misional que busca el esclarecimiento de los patrones y causas explicativas del conflicto armado interno [2], es su deber el construir herramientas que, a través del avance de la ciencia y el conocimiento, logren aportar constructivamente al entendimiento y mejoramiento de este en un marco de restauración y subsanación del conflicto armado en Colombia. Este trabajo tiene como objetivo facilitar los procesos de investigación y análisis de otras dependencias al interior de la comisión, haciendo uso del Procesamiento del Lenguaje natural (PLN). El objetivo yace en proveerles con una herramienta sencilla y ágil que les ayude a filtrar la gran cantidad de datos que sus tareas requieren analizar. Esto con el fin de posibilitarle a los investigadores corroborar su análisis, y/o llegar a conclusiones. Este trabajo centra su atención en identificar de forma automática, y dentro de un grupo de entrevistas recolectadas por la comisión, en qué construcciones lingüísticas, los testificantes podrían estar refiriendo a hechos cometidos con sevicia.

Para el caso de este informe, los análisis de preprocesamiento y procesamiento se llevaron a cabo gracias a la propuesta de prácticas concretas que puedan representar la sevicia como lo son: la decapitación, el degollamiento, el corte de franela, la castración, la incineración,

la incisión abdominal, la evisceración [3]. A través de estas prácticas, es posible encontrar indicios de sevicia respecto a construcciones lingüísticas específicas en el texto, y que pueden en este caso ser representadas por verbos. Con el fin de ampliar las posibilidades de reconocimiento dentro de un corpus dado, se procedió a una recopilación de verbos, que pudieran contener acciones cometidas con sevicia específicamente. A este respecto se usa el “léxico de la Violencia en Colombia en algunas obras de la literatura de violencia” de María Bernarda Espejo Olaya y Nancy Rozo Melo, puesto que se enfoca en el empleo de términos costumbristas en la literatura que sirven para dar cuenta de la realidad del hablante en su entorno y específicamente en términos que hacen referencia a hechos violentos en el marco del conflicto colombiano [4]. A su vez, el corpus de estudio fue el etiquetado por la Comisión de la Verdad como Masacres. Dentro del aspecto del preprocesamiento, se crea un módulo que permite al investigador analizar un espectro concreto de datos.

Por lo tanto, se obtienen resultados analizados, relacionados y graficados automáticamente de 2385 entrevistas categorizadas dentro de la categoría masacre. Los primeros resultados dan cuenta de la cantidad de eventos encontrados con posibilidad de sevicia, teniendo en cuenta cada año desde 1939. De acuerdo con análisis los periodos en los que se encontraron mayormente términos referentes a la sevicia son 1960, 1986, 1979 y 1971 siendo 1960 el año con mayor presencia de este tipo de acciones, tal y como se puede observar en la siguiente imagen (3-1).

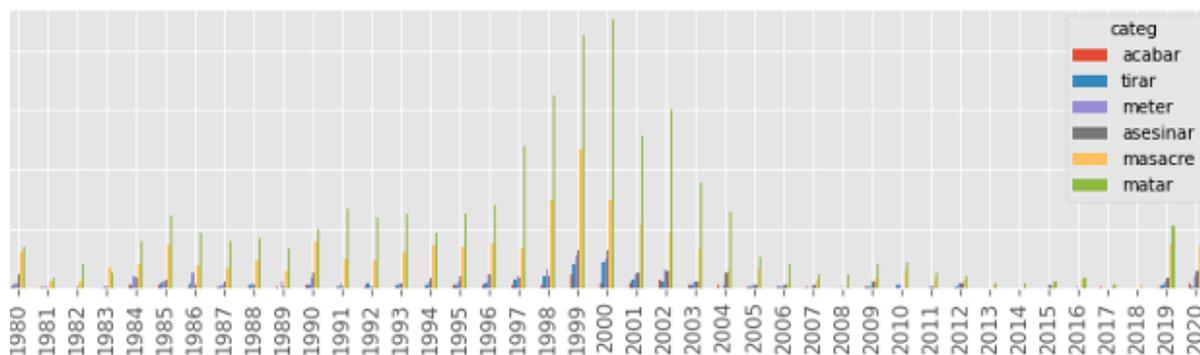


Figura 3-1: Histórico de tipos de sevicia

A continuación, los resultados referentes a dimensión de la sevicia, es decir, su gravedad respecto al año. Este procedimiento consiste en filtrar el corpus a través de los verbos y, a cada verbo filtrado asignarle un valor respecto a la Escala de sevicia que usa los números del 1 al 5. Uno, indicando un grado de sevicia que podría estar sugerido en el análisis de la construcción lingüística a la que pertenece y cinco indicando sevicia explícita en esa construcción. Del siguiente modo:

- 1,0: Verbo concomitante con hechos victimizantes, explícito: Esto quiere decir qué si bien este tipo de verbo no habla directamente sobre un acto cometido con sevicia, si in-

dican, la presencia de hechos victimizantes que son usualmente potenciales indicadores de sevicia. (Ej. agredir, atacar, disgustar).

- 2,0: Verbo con hechos victimizantes, concomitantes a la sevicia, pero implícitos: Este tipo de verbo es aquel que tienen una relación directa o cercana con actos de sevicia sin llegar a ser explícitos. (Ej. Accidentar, acabar, chocar, contender, secuestrar).

- 3,0: Verbo con hechos victimizantes, concomitante a la sevicia explícito: Este tipo de verbo es aquel que habla de hechos victimizantes que están en dirección a causar actos de sevicia, y son manifestados de manera explícita. (Ej. Abusar, cortar, humillar).

- 4,0: Verbo con presencia de sevicia implícita: Este tipo de verbo hace referencia directa a cometer un acto de sevicia, pero este queda implícito en el verbo y no manifiesta directamente el acto cometido sobre el depositario de sevicia. (Ej. Mutilar, quemar, aterrorizar).

- 5,0: Verbo con presencia de sevicia explícita: Este tipo de verbo manifiesta un acto de sevicia directo y explícito sobre el depositario de la sevicia. (Ej. Decapitar, degollar, desmembrar, eviscerar).

Así, la dimensión de sevicia es calculada en la suma ponderada de los valores de sevicia encontrados para cada palabra. Este cálculo también integra los términos repetidos, asegurando así que el grado de sevicia aumente respecto a la frecuencia de aparición de estos.

La presentación de estos resultados, donde se muestra la dimensión de la sevicia por años, se puede encontrar en la imagen **3-2**, la cual muestra al año 1999 como el año cúspide en dimensión de sevicia. Es también interesante la ascendencia en términos de violencia con presencia de sevicia desde el año 1994 y que asciende progresivamente para volver a descender, alcanzando su menor grado en el año 2008.

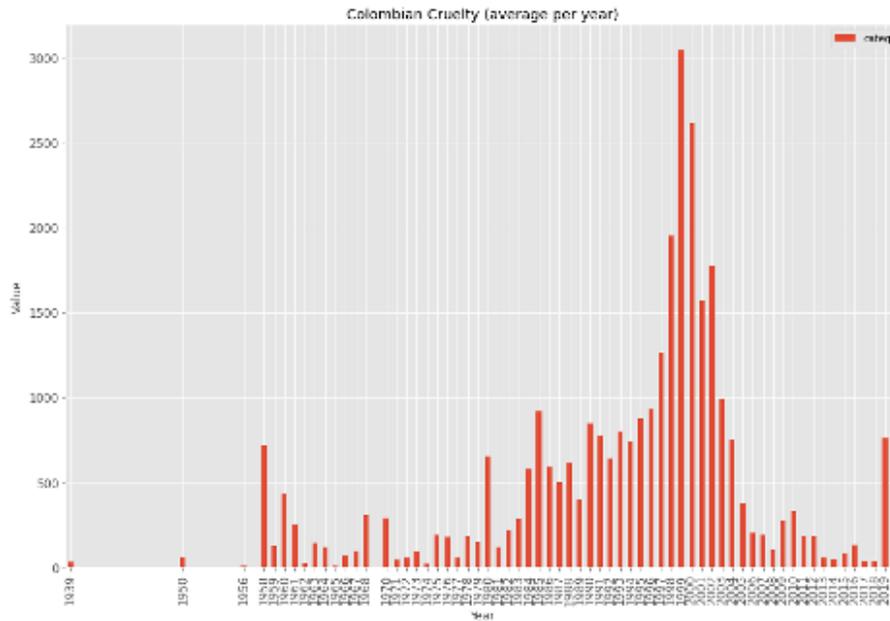


Figura 3-2: Análisis histórico dimensión de la sevicia (1939 - 2020)

Ahora bien, en relación de los promedios de sevicia encontrados dentro del corpus, se obtiene la imagen **3-3**, la cual muestra que la mayoría de las construcciones se encontraron entre las dimensiones 3 y 4, indicando que, al menos en los datos analizados, la mayoría de sevicia encontrada fue concomitante o implícita, siendo poco común la mención de sevicia explícita.

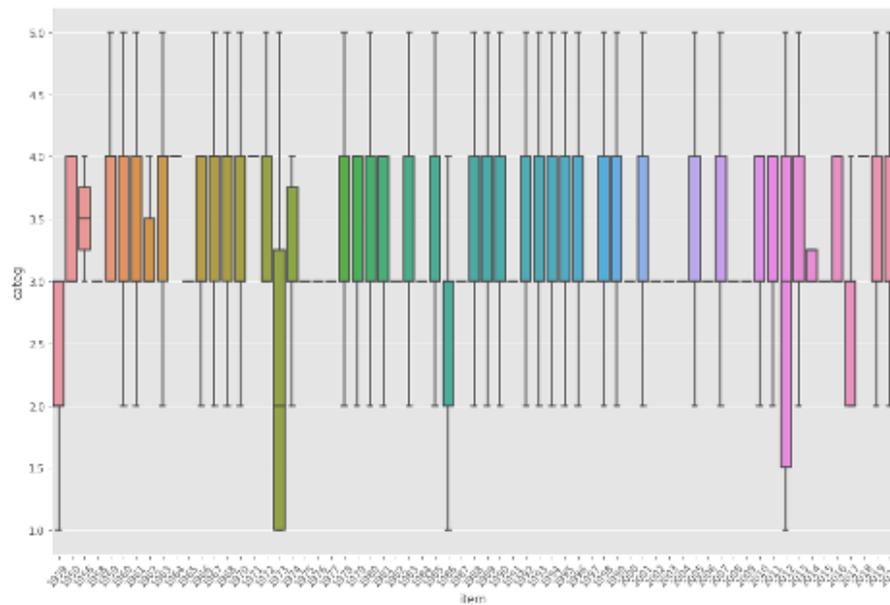


Figura 3-3: Promedio de dimensiones respecto a la muestra

No obstante, el informe no termina acá, puesto que después se procedió a realizar un nuevo

análisis en el que se tomaron tres submuestras de la muestra del corpus de Masacres con el fin de realizar un análisis detallado de los índices de sevicia que se presentan dentro de estas según los actores armados del conflicto. En este caso, las submuestras corresponden a Paramilitares, Guerrilla y Fuerza Pública. A continuación, se presentará la ejemplificación de una de estas subcategorías, cuya metodología de construcción se basó en la misma por la cual se desarrolló el informe.

La subcategoría de Fuerza Pública se toma de la submuestra de entrevistas provenientes de la muestra de la etiqueta de masacres que tenían relación con las fuerzas públicas. Como se puede apreciar en la imagen **3-4**, la cual muestra la frecuencia de términos (verbos) relacionados a la sevicia dentro de las entrevistas con respecto al año de los sucesos de esta, la frecuencia del verbo matar es el más sobresaliente y con la frecuencia más alta en los años 1958, 1998, 1999, 2000 y 2002 sobrepasando los 60 de frecuencia. De la misma manera, el término masacre es el más frecuente en los años 1959, 1970, 1978 y 1994, en este último con una frecuencia de 40.

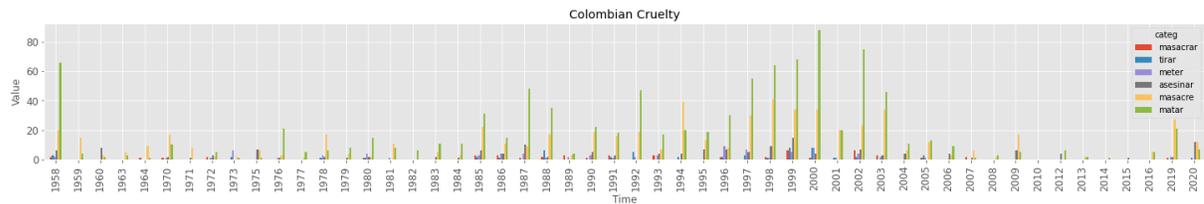


Figura 3-4: Frecuencia de los términos se sevicia en las entrevistas

Por otra parte, en la imagen **3-5** se puede identificar la ejemplificación del análisis del componente de sevicia basado en medidas descriptivas de centralización de los datos, calculando así el promedio anual del valor o índice de sevicia de todas las muestras según el lexicón. Este diagrama de barras presenta la información en dos variables referentes al año y al promedio resultante del análisis. Dentro de estos resultados se muestra que los años 1958, 1997, 1998, 1999, 2000, 2002 y 2003 presentan un promedio superior a 300, siendo 1998, 1999, 2000 y 2002 los años con un promedio más alto en cuanto a los índices de sevicia, todos superiores a 400, 1999 y 2000 cercanos al 500.

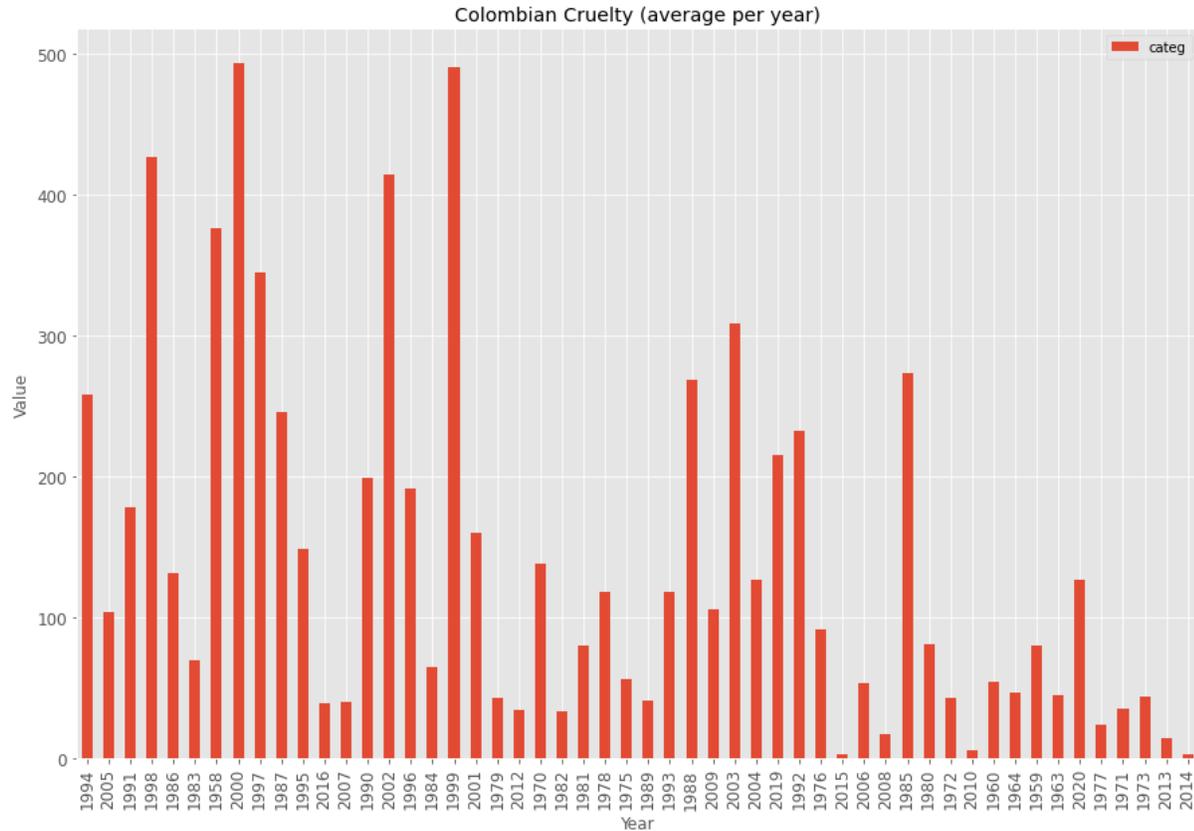


Figura 3-5: Promedio índices sevicia por año

Por último, la imagen **3-6** enseña un gráfico de tipo caja y bigotes que muestra la dispersión de los datos que puede ayudar al investigador a entender valores clave, tales como: el promedio, el percentil medio, los valores atípicos y patrones de comportamiento respecto a los años [5]. En este evidenciamos resultados muy similares a los anteriores, en el que la distribución de los datos de la mayoría de los años se localiza entre los niveles 3 y 4 de sevicia, siendo en 1985 y 2013 donde se ve un promedio de esta distribución en 3.5. De la misma forma, en el apartado de Fuerza Pública vuelve a resaltar el año 1973 por su distribución de datos que se concentra entre 1 y 4 y tiene como promedio el nivel 2 de sevicia.

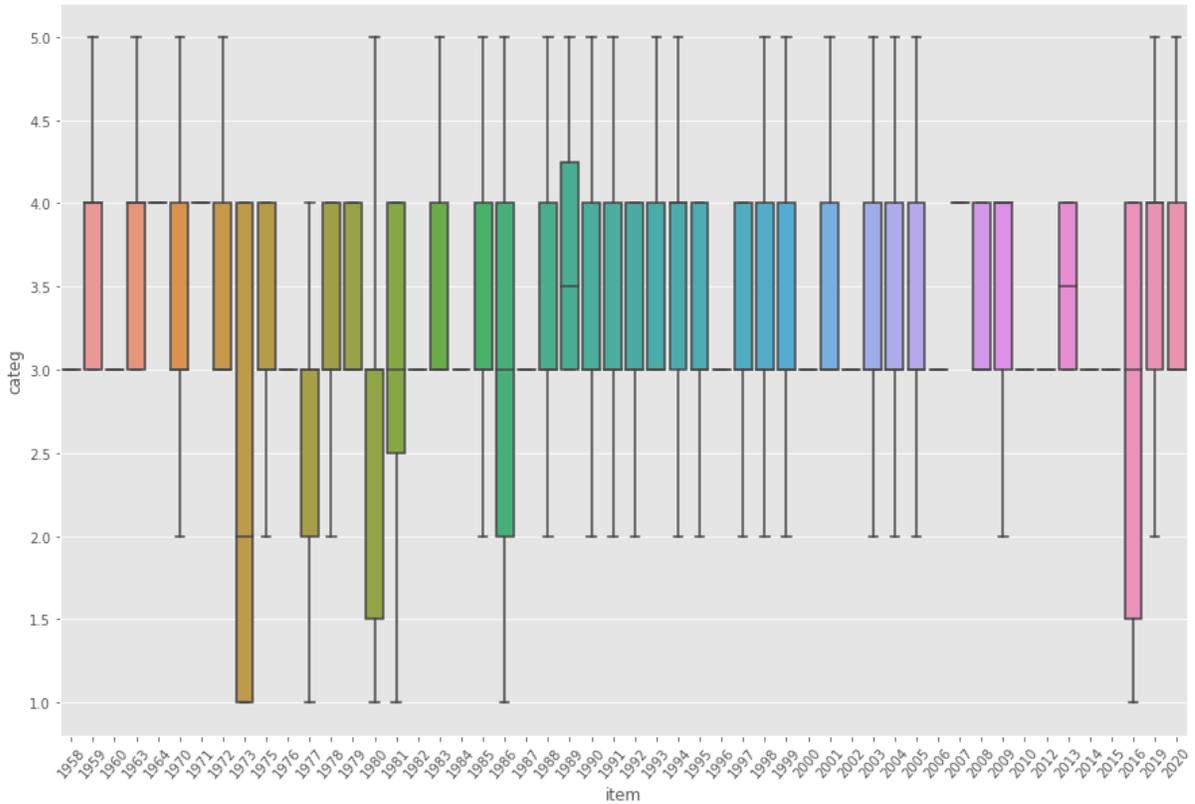


Figura 3-6: Caja y Bigotes: Fuerza pública

Como conclusiones generales, es necesario aclarar que estos son resultados preliminares correspondientes únicamente a la fase 1 de desarrollo del programa anteriormente descrito. Esto quiere decir, que este es apenas un indicador de trabajo, y más bien una ruta de dirección al perfeccionamiento de los análisis posibles con la máquina desarrollada. Algunos de los obstáculos que se enfrentaron fueron sobre todo el lograr una normalización de los datos, es decir construir unas reglas y parámetros suficientemente estables, con el fin de minimizar la redundancia de datos, así facilitando su gestión [6]. Por otro lado, el fenómeno de la lengua es complejo, y su estudio objetivo es una tarea compleja que conlleva un buen tiempo de desarrollo. Finalmente, también es necesario recordar que la herramienta no pretende indagar en la verdad o falsedad de los datos consultados, sino que es más bien una herramienta metodológica para el análisis.

Bibliografía

- [1] Oxford, “Oxford languages,” 2021.
- [2] Comisión, “¿qué es la comisión de la verdad?,” 2021.
- [3] A. F. Suárez, “La sevicia en las masacres de la guerra colombiana,” *Análisis político*, vol. 21, no. 63, pp. 59–77, 2008.
- [4] M. B. E. Olaya and N. R. Melo, “El léxico de la violencia en Colombia en algunas obras de la literatura de violencia,” in *Congreso Nacional y Internacional de Lingüística, Literatura y Semiótica*, 2015.
- [5] T. D. V. Catalogue’, “Caja y bigotes,” 2021.
- [6] R. Elmasri, S. B. Navathe, and G. E. Roberto, *Sistemas de bases de datos: Conceptos fundamentales*. Addison-Wesley, 1989.

4 Clasificadores, intensidades y lexicones

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las entrevistas de los diferentes corpus utilizados dentro de los diversos informes que comprenden múltiples objetivos y preguntas de investigación.

El trabajo metodológico se comprende por medio del procesamiento de entrevistas pertenecientes al corpus recolectado por la Comisión de la Verdad bajo su mandato de acuerdo con los diversos intereses de los diferentes informes. Los corpus procesados contienen varios números de entrevistas, no obstante, la metodología continúa siendo la misma. Cuando se tokenizan las entrevistas de la plataforma para su análisis, se encuentran dentro de ellas distintos términos irrelevantes denominados “stopwords”, los cuales se limpian de los resultados parciales, así como los caracteres especiales como los paréntesis, tildes, signos de puntuación, etc. Una vez son filtrados los resultados, se procesan a través de una herramienta de conteo utilizando las librerías Spacy y Matplotlib, que se explicarán a continuación, arrojando una categorización de cada término del corpus según los miembros de las parejas de nodos de interés.

Spacy es una librería de software para procesamiento de lenguajes naturales desarrollado por Matt Honnibal y programado en lenguaje Python. Su utilidad en nuestra propuesta es la capacidad que otorga para tokenizar, es decir, dividir el texto en unidades lingüísticas individuales de manera automática [1]. De esta manera, Spacy nos permite reconocer la categoría gramatical de una palabra dada, además de filtrar su lema, lo que hace más fácil la identificación sin importar la flexión verbal en la que se encuentre.

Matplotlib es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python. Fue utilizada para proporcionar resultados gráficos de los datos analizados. Estos gráficos son algunos de barras, caja y bigotes y nubes de palabras.

Para cada uno de los informes se empleó un lexicón específico, puesto que es la base fundamental en la construcción de sistemas computacionales que posibiliten la interacción entre la máquina y el hombre [2]. Es decir que, un lexicón posibilita la automatización de los procesos lingüísticos como la comprensión, producción o adquisición de una lengua [2]. Una vez desarrollado el lexicón, este se utiliza para la generación de categorías de estudio depen-

dientes del objetivo o enfoque de cada trabajo. Con estas categorías definidas se procede a clasificarlas por grados de intensidad de 0 a 5, puesto que algunos términos se relacionan con más de una categoría.

4.1. Ejecuciones extrajudiciales en el conflicto armado colombiano

Este informe fue culminado el 27 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural.

Como parte de los hechos victimizantes dados durante el conflicto armado colombiano se encuentra el tema de las ejecuciones extrajudiciales. Con positivos se entendían las bajas en combate, por las que los militares o soldados recibían bonificaciones monetarias u otros beneficios como ascensos y permisos. Por lo mismo, cada baja se sumaba a los beneficios que recibían estos uniformados y que a fin de cuentas eran casos de ejecuciones extrajudiciales o desapariciones forzadas que causaban inmensurable sufrimiento en los familiares de las víctimas y que no recibían el debido proceso de investigación. Es en estos hechos victimizantes en los que se pretende profundizar en el presente informe. En este se enfocará el estudio en las ejecuciones extrajudiciales en el corpus de entrevistas recolectado por la Comisión de la Verdad. Con esto en mente, se pretende entender de manera más cercana la percepción que tienen las víctimas y testigos del conflicto armado de las ejecuciones extrajudiciales. ¿Cómo se dieron? ¿Por qué razones? ¿Quiénes? ¿Cómo se distorsionó la figura de las víctimas? ¿Cómo se manejaron las investigaciones? Todo esto usando como base el corpus de entrevistas recolectado por la Comisión de la Verdad y con el fin de esclarecer un poco los sucesos y condiciones que rodearon a las ejecuciones extrajudiciales y las consecuencias de estas.

A partir del árbol de etiquetas en el que la Comisión organiza las entrevistas se seleccionaron aquellas dentro del numeral 14. Dominio Transversal, 2. Hechos victimizantes, que a su vez contiene la etiqueta de Homicidios y dentro de esta se seleccionaron las entrevistas pertenecientes a 1. Ejecución extrajudicial, sumaria o arbitraria y 2. Ejecución extrajudicial presentada como muerte en combate o falso positivo. Estas entrevistas presentan los relatos relacionados con los testimonios de personas cuyos familiares, conocidos o amigos fueron víctimas de las ejecuciones extrajudiciales o falsos positivos. A través del procesamiento del corpus, se encontró un resultado promedio de 298.9 tokens aproximadamente con una diversidad lexical promedio de 0.5 aproximadamente.

A su vez, para complementar el análisis hallado en las entrevistas, se elaboró un lexicón basado en ejemplos previos como el construido por Marguerite Feitlowitz en su libro A

Lexicon of Terror (2011), en el cual analiza el lenguaje utilizado durante la dictadura Argentina resaltando términos como desaparecidos, subversión, entre otros; que contribuyeron a la construcción de un enemigo común dando paso a una serie de torturas y ejecuciones extrajudiciales brutales e inhumanas. De la misma manera, se recopilaron términos asociados a las ejecuciones extrajudiciales provenientes de la definición de estas o los eufemismos usados para cubrirlas como el ya mencionado fuego cruzado. Estos términos fueron posteriormente categorizados para así ser comparados con el corpus de entrevistas previamente seleccionado y encontrar diversas relaciones que nos permitan tener una idea más clara de cómo se ve la figura de las ejecuciones extrajudiciales en el corpus de entrevistas que cuentan los testimonios de víctimas o participantes del conflicto armado colombiano.

En consecuencia, para el estudio de las entrevistas referentes a las ejecuciones extrajudiciales se crearon cuatro categorías que buscan clasificar los términos según su relación con los actores de estos hechos o con las formas en las que estos hechos se llevaron a cabo, las categorías son las siguientes:

- Víctimas: Esta categoría hace referencia a aquellas palabras o términos que tienen relación, afectan o describen de alguna manera a las víctimas de las ejecuciones extrajudiciales, respondiendo a la pregunta ¿Quién? O ¿A quién?
- Victimarios: En esta categoría se hace referencia a las personas que Todorov menciona como malhechores y que son responsables o tienen algún tipo de relación con el crimen de las ejecuciones extrajudiciales.
- Testigos: En esta categoría se encuentran todas aquellas palabras referentes a personas responsables del relato, quienes están relacionados con los hechos como familiares, amigos o conocidos de las víctimas; o bien quienes presenciaron o supieron algo con respecto a los hechos.
- Métodos: En la última categoría se posicionan los términos o palabras referentes a procedimientos o modalidades que se llevaban a cabo para cometer dichas ejecuciones.

La posterior clasificación por grados de intensidad es la siguiente:

- 0: No presenta relación
- 1: Presenta relación no directa por consecuencia: Se relaciona en cuanto la categoría se ve levemente afectada como consecuencia del término
- 2: Presenta relación no directa: Se relaciona levemente a la categoría por sí mismo.
- 3: Presenta relación directa por consecuencia: Se relaciona o afecta a la categoría como consecuencia del término.

- 4: Presenta relación directa: Se relaciona fuertemente con la categoría.
- 5: Autor o recipiente directo del término: Es quien comete la acción, representa o recibe el término.

Para la presentación de los resultados, primero se exponen las frecuencias de algunos términos en una gráfica de barras acompañada por una nube de palabras. Un ejemplo de esto se encuentra en la imagen 4-1 donde se presentan los verbos más frecuentes dentro del corpus, los cuales son decir, matar, tener, hacer, saber, entre otros. Esta información es representada de manera diferente en la imagen 4-2, dónde se presenta la nube de palabras de todos estos verbos en diferentes colores. Entre mayor sea el tamaño de las palabras, mayor presencia tiene dentro del corpus.

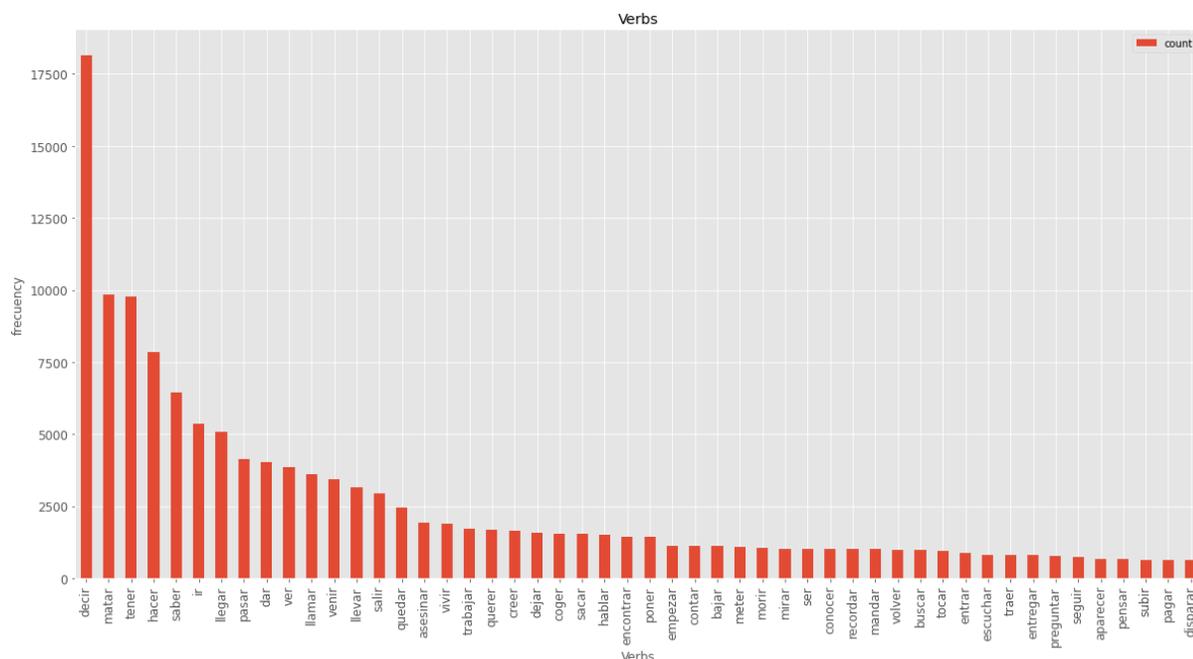


Figura 4-1: Verbos más frecuentes



Figura 4-2: Nube de palabras: verbos más frecuentes

A su vez, es importante declarar que los resultados también se presentan con ayuda de modelos fastText, el cual es un recurso abierto, liviano y gratuito, que permite a los usuarios aprender sobre representaciones y clasificadores de texto [3]. Al procesar el corpus, este se filtró bajo la etiqueta de Homicidios, en la cual se encuentran otras etiquetas de tercer nivel que corresponden a ejecuciones extrajudiciales: 1. Ejecución extrajudicial arbitraria; 2. Ejecución extrajudicial falso positivo. El objetivo de dicho proceso es mostrar la dispersión de las entrevistas etiquetadas de acuerdo con las similitudes lexicales que entre estas se presenten. De esta manera, a cada entrevista se le asigna un pequeño punto con un color específico que la identifica como parte de x etiqueta. Un ejemplo de esto puede ser percibido en la imagen 4-3 donde se observa que las entrevistas se agrupan de acuerdo con las sub-etiquetas que las identifican, si bien hay algunas dispersas y alejadas o entre mezcladas, lo más evidente son los grupos que se forman. En el grupo más grande podemos observar en color magenta las entrevistas de la etiqueta Hechos- Homicidios, esta agrupación evidencia que dentro de dichas entrevistas la variedad lexical o la forma de narrar los hechos es similar entre ellas y no tan parecida a las demás etiquetas. Otro grupo relevante dentro de la gráfica es aquel representado por el color amarillo que agrupa las entrevistas de Hechos- Homicidios- Masacre y del que también es posible observar un grupo más pequeño cercano al magenta.

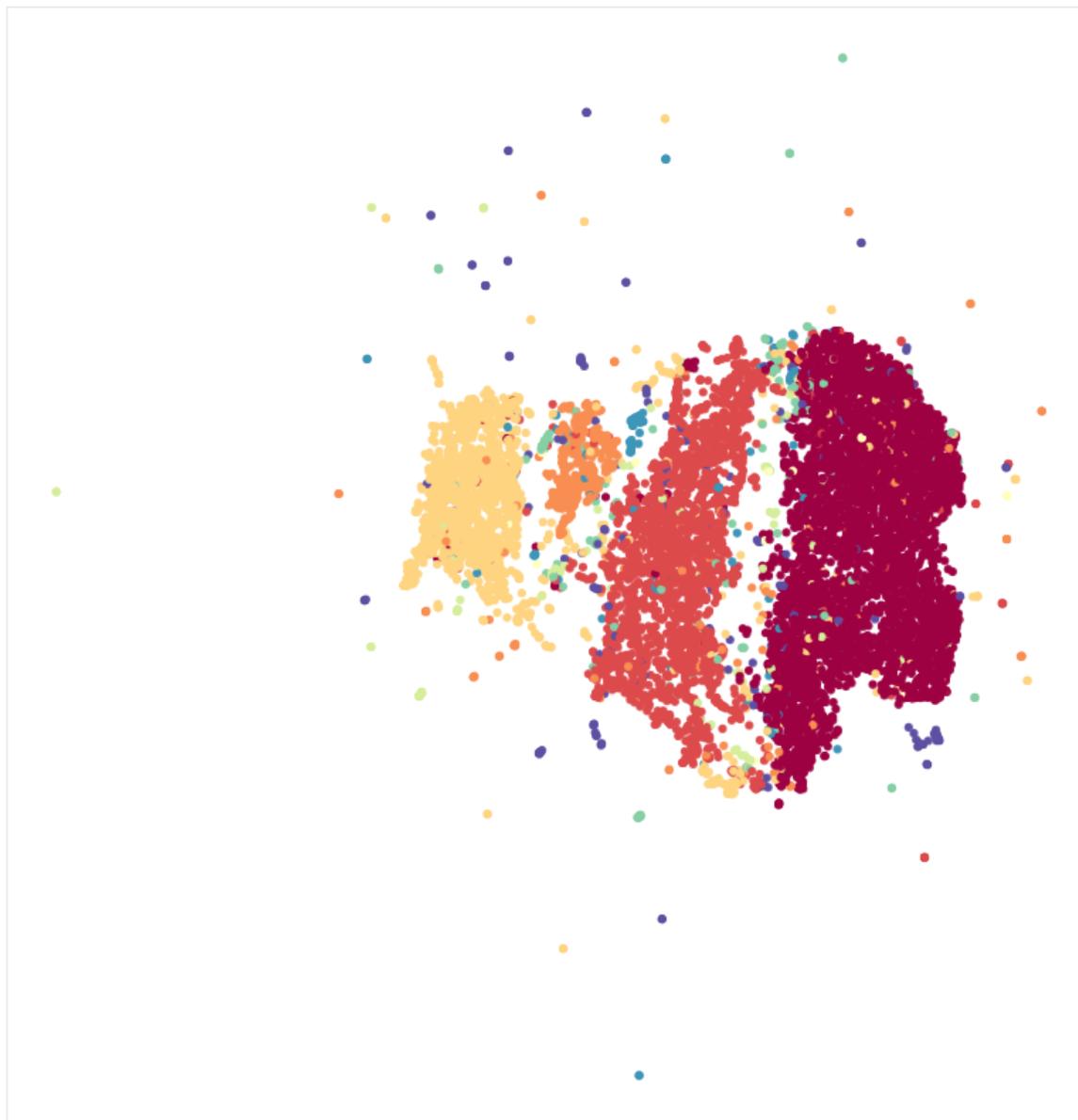


Figura 4-3: FastText: Homicidios

Por último, para la presentación de resultados de las categorías desarrolladas para este informe en específico, se obtiene la imagen 4-4, la cual muestra el análisis del componente con respecto a la categoría de víctimas basada en medidas descriptivas de centralización de los datos, calculando así el promedio anual del valor o índice de los términos ligados a la categoría de víctimas de toda la muestra según el lexicon de acuerdo con el año de los hechos. Por lo mismo, es posible evidenciar que los índices de términos con referencia a las víctimas más altos se presentan en los años 2002 y 2000, donde se supera el promedio de 16000 o se llega a él. Esto son seguidos por los años 2001 y 2005 donde se muestran valores cercanos a 10000. Esto quiere decir que estos son los años en los que se narran más hechos o se mencionan

con más frecuencia a las víctimas como protagonistas del testimonio. Esto se relaciona con el aumento de estos hechos victimizantes que menciona la Amnistía Internacional y también con que entre el año 1996 y 2002 “el conflicto armado alcanzó su nivel más crítico como consecuencia del fortalecimiento militar de las guerrillas, la expansión nacional de los grupos paramilitares, la crisis del Estado, la crisis económica y la reconfiguración del narcotráfico” [4, p. 167]. Lo que permitió que delitos como los secuestros, las masacres y las desapariciones forzadas se incrementaran, y muchas de esas desapariciones serían después confirmadas como ejecuciones extrajudiciales o reclutamiento forzado.

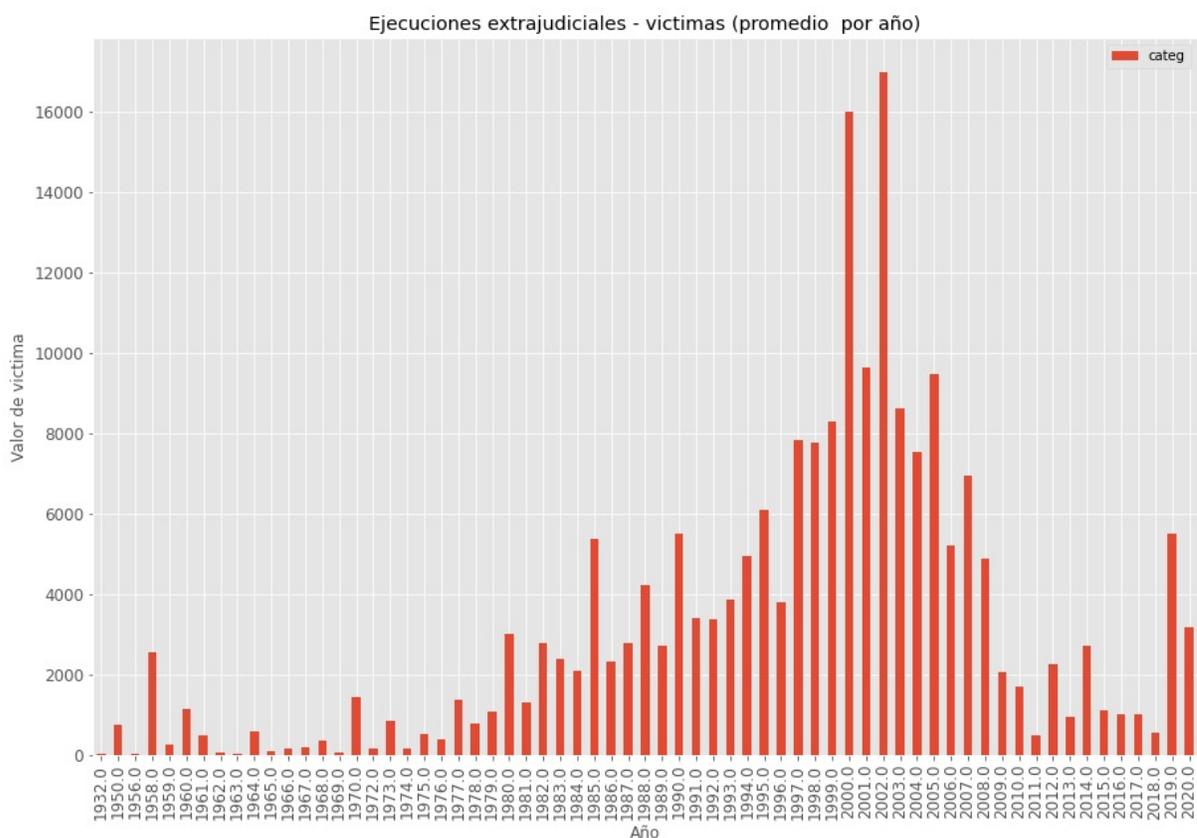


Figura 4-4: Ejecuciones Extrajudiciales: Valores referentes a víctimas

Otra forma de la representación de los datos de esta categoría en específico es a través de la gráfica de caja de bigotes vista en la imagen 4-5, dónde se muestra la dispersión de los datos que puede ayudar al investigador a entender valores clave, tales como: el promedio, el percentil medio, los valores atípicos y patrones de comportamiento respecto a los años [5]. En esta se presenta que en la categoría de Víctimas entre los años 1932 y 2020 la intensidad de los términos encontrados en el corpus con base en el lexicón muestra que el promedio de intensidades en su mayoría se encuentra entre el nivel 1 y 5 o 1 y 3. Lo que quiere decir que la mayoría de los términos encontrados están indirectamente relacionados por consecuencia con la categoría de víctima o en el caso del nivel 3 directamente relacionado por consecuencia. En

este resalta el año 1956 la mediana, representada por las líneas en los cuadros de promedio, se encuentra en el nivel 4, es decir, que los términos presentan una relación directa. Otro patrón importante presentado en la gráfica está entre los años 1981 y 2005, cuya mediana está en el nivel 3 es decir relación directa por consecuencia con las víctimas.

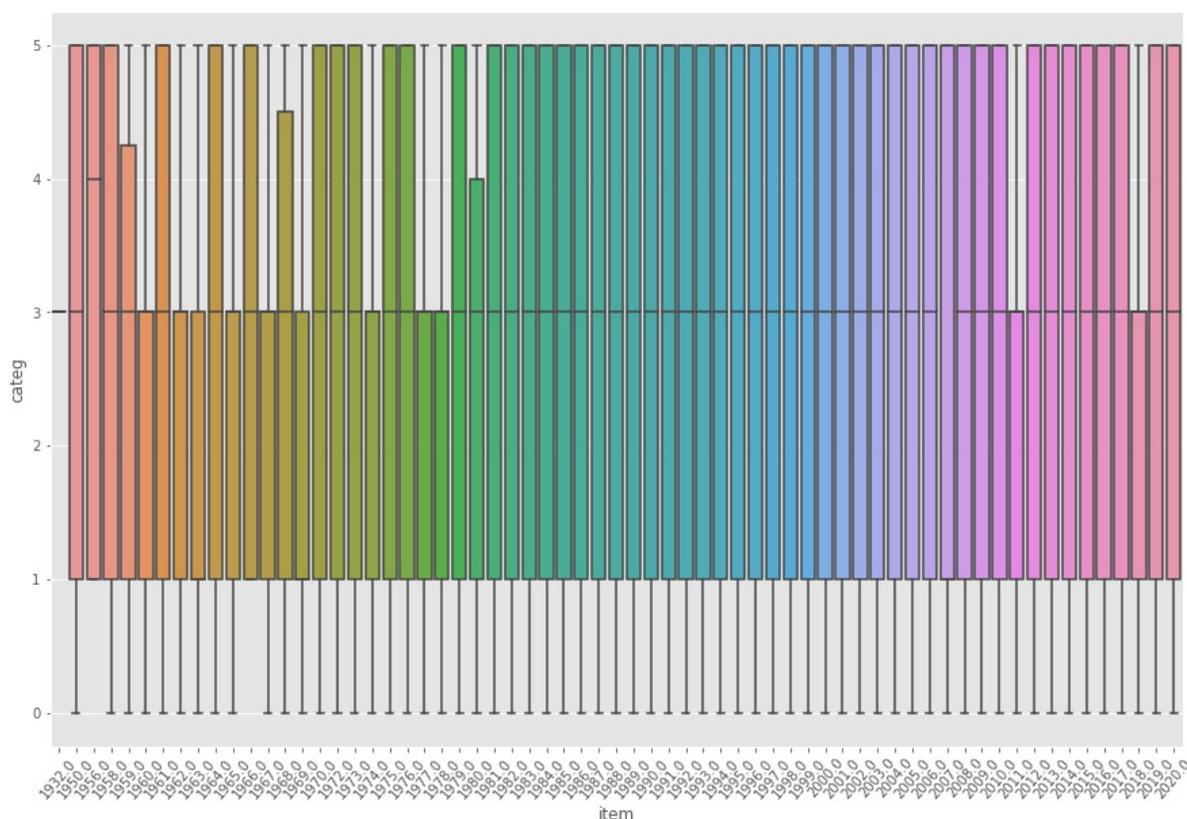


Figura 4-5: Ejecuciones Extrajudiciales: Caja y Bigotes Víctima

Como conclusiones generales, es necesario resaltar las coincidencias históricas con el análisis de corpus. Es importante volver a la premisa objeto del estudio. Las ejecuciones extrajudiciales que son una violación a los Derechos Humanos y que desafortunadamente resaltan en la historia del país y sobre todo en el conflicto armado. Existen dos momentos relevantes en cuanto a estas ejecuciones o falsos positivos. El primero de ellos se da durante los años 2000, en donde se presenta un incremento significativo a los casos de ejecuciones extrajudiciales y que la Amnistía Internacional liga directamente a la instauración del Plan Colombia durante el 2000. Además de que incrementó las violaciones a los Derechos Humanos, los cultivos de coca, el número de víctimas del conflicto armado interno, la minería ilegal y la idea de la fuerza pública como mercenarios.

Por lo mismo, se evidencia en los análisis realizados que la información recolectada en el corpus de entrevistas de la Comisión de la Verdad con respecto a estos sucesos, contienen

mayores frecuencias y valores en las categorías presentadas por el lexicón en los años 2000, 2002, 2005, etc. que corresponden con las fechas resaltadas anteriormente. Adicionalmente, es relevante resaltar a la categoría Victimarios que es la que presenta mayores valores con respecto a las intensidades de los términos presentados en el lexicón. Lo que podría indicar un protagonismo o alta mención de estos, o de términos relacionados con ellos, en los relatos de los entrevistados.

De igual manera, el escándalo de los falsos positivos del año 2006 derivado de la implementación de la Política de Seguridad Democrática coincide con los altos valores y frecuencias del año 2005. Esto podría explicar la importancia de los victimarios en las narrativas y la principal diferencia de este hecho con respecto a la categoría de homicidio.

Por último, se podrían justificar los altos valores de los términos entre el año 2000 y 2010, así como la presencia de palabras como jóvenes o miembros de la familia en el corpus. Frecuencias que se evidencian en los resultados del análisis gramatical presentado anteriormente. Así mismo, la similitud de valores entre la categoría de testigo y de métodos en los relatos podría deberse a la necesidad que este hecho victimizante tiene de consultar a terceros y de describir las acciones que se llevaron a cabo por los victimarios, quienes por su naturaleza específica poseen un poco más de protagonismo en los testimonios.

4.2. Duelo en los familiares de desaparecidos en Colombia

Este informe fue culminado el 28 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Este trabajo tiene como propósito indagar cómo los familiares de personas desaparecidas en Colombia, producto del conflicto armado, se ven atravesados por el duelo; es decir, observar la manera en la que viven y gestionan el duelo. Estas personas serán claves a la hora de observar la intensidad del duelo, la forma cómo se habla en este tipo de duelo, desde la búsqueda o el hallazgo, y los tipos de emociones y actos que se presentan en los familiares de los desaparecidos. En esta ocasión, se tendrán en cuenta factores como la posesión del cuerpo de la víctima, la falta del cuerpo de la víctima, el conocimiento sobre el paradero de la víctima y el total desconocimiento sobre el paradero de la víctima. Por lo tanto, es necesario examinar el fenómeno de la violencia desde la perspectiva, el lenguaje y las prácticas de los sufrientes, ya que son ellos quienes nos muestran la manera en la que negocian y obtienen reductos de dignidad, resisten y reconstruyen sus relaciones cotidianas, y sobrellevan la huella de la violencia.

A partir del árbol de etiquetas se seleccionaron aquellas bajo el nombre de desapariciones forzadas. Esta etiqueta tiene sub-etiquetas: 1) Desaparición selectiva; 2) Desaparición encubridora; 3) Desaparición como medio de intimidación; 4) Fosas comunes. Con la elección de estas entrevistas se busca hallar las palabras más relevantes para el análisis del duelo en los

familiares de desaparecidos en Colombia. De la misma manera, se espera encontrar palabras claves en términos de búsqueda y hallazgo de desaparecidos en Colombia para entender la manera en la que los familiares afrontan el duelo y buscan respuestas. En otras palabras, se esperan encontrar palabras que ayuden a identificar la manera en la que las víctimas atraviesan y viven el duelo ante la pérdida de un familiar. Las entrevistas de desaparición forzada tienen la característica de presentar los relatos de las víctimas o de testigos de los hechos cuyos familiares y/o seres queridos fueron detenidos por grupos armados y posteriormente desaparecidos. Cabe mencionar, que un aspecto relevante encontrado al realizar la lectura de algunas de estas entrevistas era el intento de búsqueda por algún familiar y en algunos casos, muy pocos, el hallazgo de los restos de un ser querido. A través del procesamiento del corpus, se encontró un resultado promedio de 271.4 tokens aproximadamente.

Dentro de este trabajo se realizó la construcción de un lexicón del duelo. Este lexicón permitirá analizar las entrevistas correspondientes a la etiqueta de desaparición forzada mediante dos categorías:

- Categoría de búsqueda: En esta categoría se incluirán términos que indiquen o apunten la actividad de los familiares de los desaparecidos en Colombia buscando a sus seres queridos incluyendo el dolor y angustia que esto representa.
- Categoría de hallazgos: En esta categoría se incluirán términos que indiquen o apunten el suceso de un cuerpo recuperado por parte de los familiares de los desaparecidos en Colombia.

Con estas dos categorías se busca identificar como negocian con el dolor los familiares de los detenidos. Es decir, responder las siguientes preguntas qué hacen cuando desaparece un familiar (si lo buscan, si su vida continua igual), cómo viven el proceso del duelo (lo pueden hacer o no), y sobre todo qué sucede cuando un ser querido ya no está.

La posterior clasificación por grados de intensidad es la siguiente:

- 0: Nada relevante: Palabras que no aplican para una categoría específica de búsqueda o hallazgo. (Ej. Vacío, Normalidad, Emoción).
- 1: Muy poco relevante: Palabras que por sí solas no tienen relación evidente con el tema del duelo. Sin embargo, resultan útiles dentro de los análisis. (Ej. Cuerpo, Rostro, Tiempo).
- 2: Poco relevante: Palabras que, aunque tienen que ver con el tema del duelo, no son específicas del tema y pueden corresponder a otros similares como muerte o emociones en general. Sin embargo, resultan útiles para analizar de lo general a lo específico. (Ej. Limbo, Llanto, Culpa).

- 3: Medianamente relevante: Palabras que denotan duelo de forma leve. Están asociadas a las emociones que se producen durante este proceso y de manera especial se asocian con la ausencia y muerte de un ser querido. Además, aparecen como los primeros “síntomas” o reacciones del duelo. (Ej. Dolor, Recuerdo, Memoria).
- 4: Relevante: Palabras asociadas fuertemente con el tema del duelo. Son manifestaciones del duelo en emociones o acciones. (Ej. Ausencia, Despojos, Pérdida).
- 5: Muy relevante: Palabras específicamente relacionadas con el tema del duelo. En especial aquellas que buscan responder a la pregunta de investigación respecto a la intensidad del duelo en las etapas de búsqueda de un cuerpo o hallazgo de este. (Ej. Fosa Común, Desconsuelo, Ubicar, Hallar, Encontrar).

Para la presentación de los resultados, primero se exponen las frecuencias de algunos términos en una gráfica de barras acompañada por una nube de palabras. Un ejemplo de esto se encuentra en la imagen 4-6 donde se presentan los adjetivos más frecuentes dentro del corpus, los cuales son bueno, desaparecido, paramilitar, muerto, primero, entre otros. Esta información es representada de manera diferente en la imagen 4-7, dónde se presenta la nube de palabras de todos estos verbos en diferentes colores. Entre mayor sea el tamaño de las palabras, mayor presencia tiene dentro del corpus.

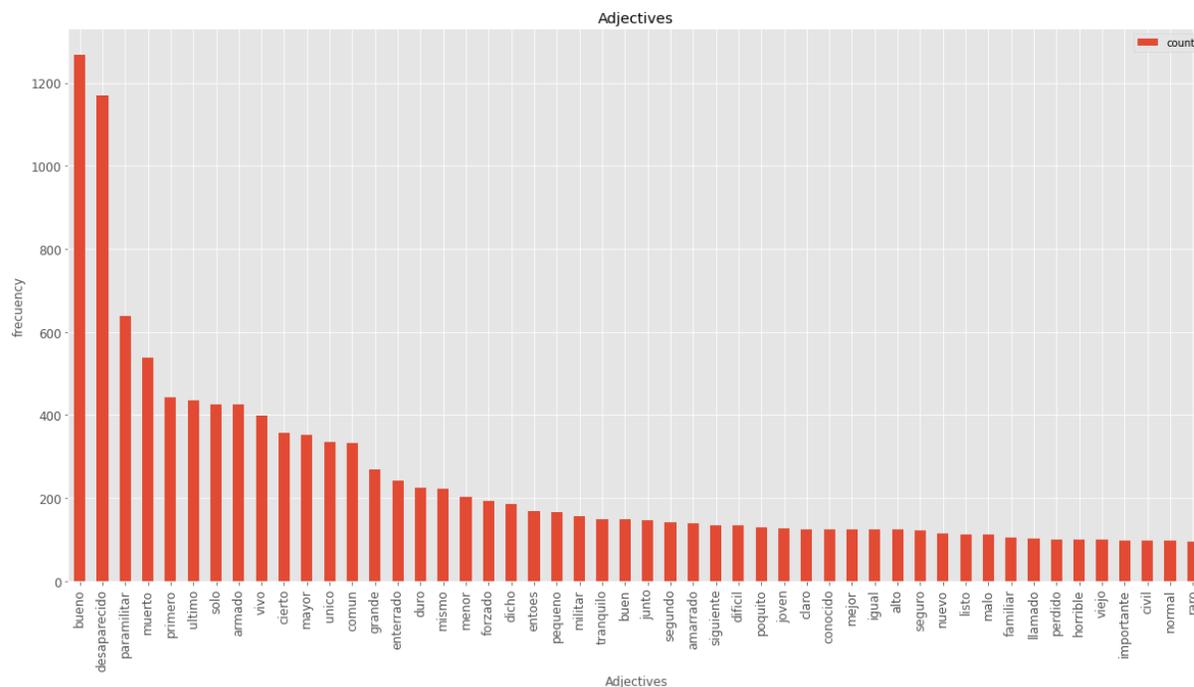


Figura 4-6: Adjetivos más frecuentes

y también a la manera en que se usan en sus contextos [3].

Ahora bien, el corpus de entrevistas pertenecientes a la etiqueta de desapariciones forzadas fue procesado con el software de fasttext. Lo que se busca es presentar la dispersión de las entrevistas etiquetadas de acuerdo con las similitudes lexicales que tengan entre sí. Dentro de las etiquetas de desapariciones forzadas también se encuentran las etiquetas de Desaparición selectiva, Desaparición encubridora, Desaparición como medio de intimidación, Fosas comunes. Así, a cada entrevista se le asigna un pequeño punto con un color específico que la identifica como parte de x etiqueta. Un ejemplo de esto puede ser percibido en la imagen 4-8 donde se pueden observar 3 grupos claramente definidos. El de color magenta corresponde a las entrevistas bajo la etiqueta de Desaparición Forzada; el color verde claro responde a las etiquetas de Desaparición Selectiva; y el color morado corresponde a la etiqueta de Fosas Comunes. De la misma manera, aunque no agrupadas en un grupo considerable aparecen las entrevistas de Desaparición Encubridora de color naranja. Estas entrevistas comparten similitudes lexicales entre sí mismas al estar agrupadas en grupos específicos. El grupo más grande de entrevistas corresponde al de desaparición forzada y este, a su vez, también comparte algunas similitudes con las etiquetas de fosas comunes, desaparición encubridora y desaparición selectiva como se muestra a continuación.

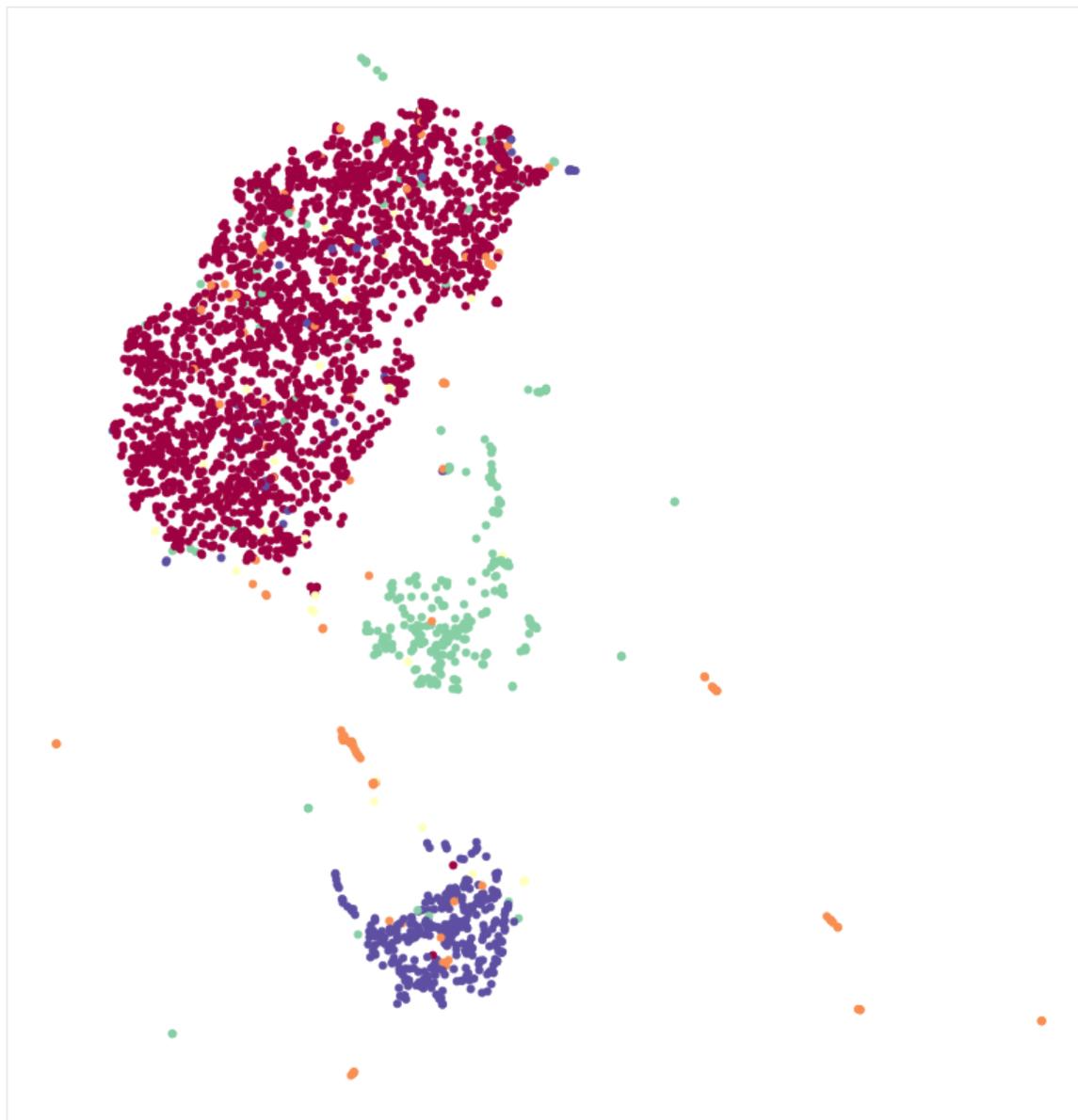


Figura 4-8: Fasttext: Desapariciones Forzadas

Por último, para la presentación de resultados de las categorías desarrolladas para este informe en específico, se obtiene la imagen 4-9, la cual muestra el análisis del componente con respecto a la categoría del análisis histórico de la dimensión del duelo para el caso de la categoría de Búsqueda. Gracias a esto, para empezar, resulta ser el año 2000 el que mayor intensidad en duelo posee con respecto a los términos del lexicon de duelo, siendo así el año cúspide en dimensión del duelo con una intensidad de un poco más de 3000. Lo anterior puede deberse a que entre el año 1996 y 2002 “el conflicto armado alcanzó su nivel más crítico como consecuencia del fortalecimiento militar de las guerrillas, la expansión nacional de los grupos paramilitares, la crisis del Estado, la crisis económica y la reconfiguración del

narcotráfico” [4, p. 33]. De esta manera, crímenes de lesa humanidad como los secuestros, las masacres y las desapariciones forzadas tuvieron un periodo de esplendor. Por ejemplo, según el informe del Grupo [4], el Catatumbo, en el Norte de Santander, se convirtió en un escenario de acciones violentas como desapariciones forzadas entre 1999 y el 2000 (pp. 171-172). Otro aspecto importante es que la desaparición forzada fue tipificada como un delito solo “a partir de la Ley 589 del 2000” [4, p. 58], debido a la alta afluencia de víctimas como los familiares de los desaparecidos. Del mismo modo, los años 2002 con una intensidad de 3000, 2001 con una intensidad de alrededor de 2400, y 1999 con una intensidad de 2300 también entran dentro del periodo de agudización del conflicto armado. Motivo por el cual la incidencia de víctimas que perdieron a sus seres queridos puede ser más alta a la vez que el duelo después de la muerte y la desaparición aparece de forma representativa en estos años. Así, la intensidad del duelo en estos años específicos puede deberse a la concurrencia de varias palabras con cierta intensidad o por la concurrencia de algunas palabras con un valor de intensidad muy alto.

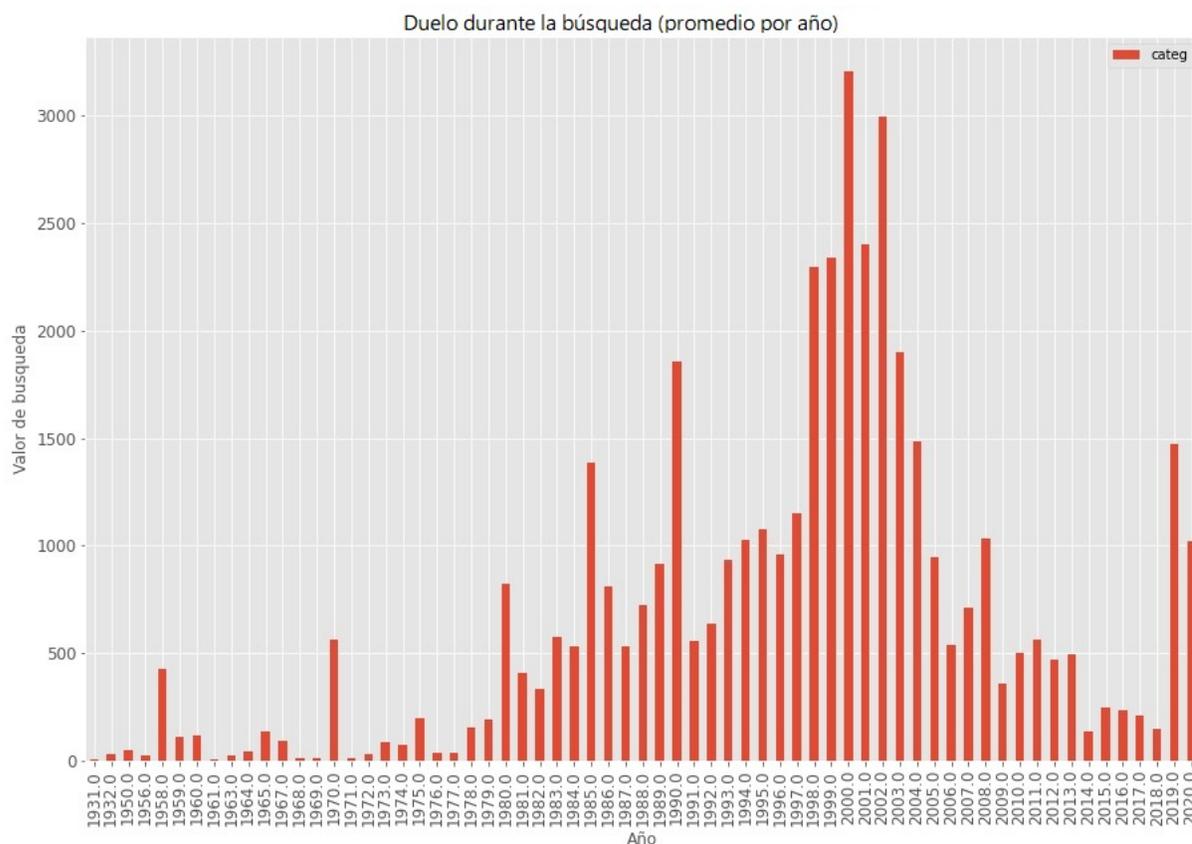


Figura 4-9: Análisis histórico dimensión del duelo (1931 – 2019) Categoría de Búsqueda

Otra forma de la representación de los datos de esta categoría en específico es a través de la gráfica de caja de bigotes vista en la imagen 4-10, dónde se muestra la dispersión de los datos que puede ayudar al investigador a entender valores clave, tales como: el promedio,

el percentil medio, los valores atípicos y patrones de comportamiento respecto a los años. En esta se pretende indagar sobre el tipo de acciones que emprenden los familiares de los desaparecidos una vez no cuentan con información sobre su ser querido y la manera como puede ser parte del duelo. En este caso, de 1931 al 2020, los datos van desde la categoría de intensidad 0 (muy poco relevante) hasta 5 (muy relevante). Sin embargo, no hay muchas palabras dentro de la categoría entre 0 y 1. La concentración de palabras se evidencia entre las categorías de intensidades 1 a 3 por lo que en su mayoría las palabras que indican duelo para esta categoría están dentro de esas intensidades. Por último, en años como 1959, 1975, 1981, 1984, 1988, 2014 y 2019 el índice de duelo alcanza la categoría de intensidad 5 es decir muy relevante. Esto podría sugerir que durante estos años hubo eventos que ocasionaron la pérdida de un familiar por desaparición forzada y la consecuente indagación por más información. Cabe señalar que los espacios en blanco refieren a falta de información y/o palabras para esos años.

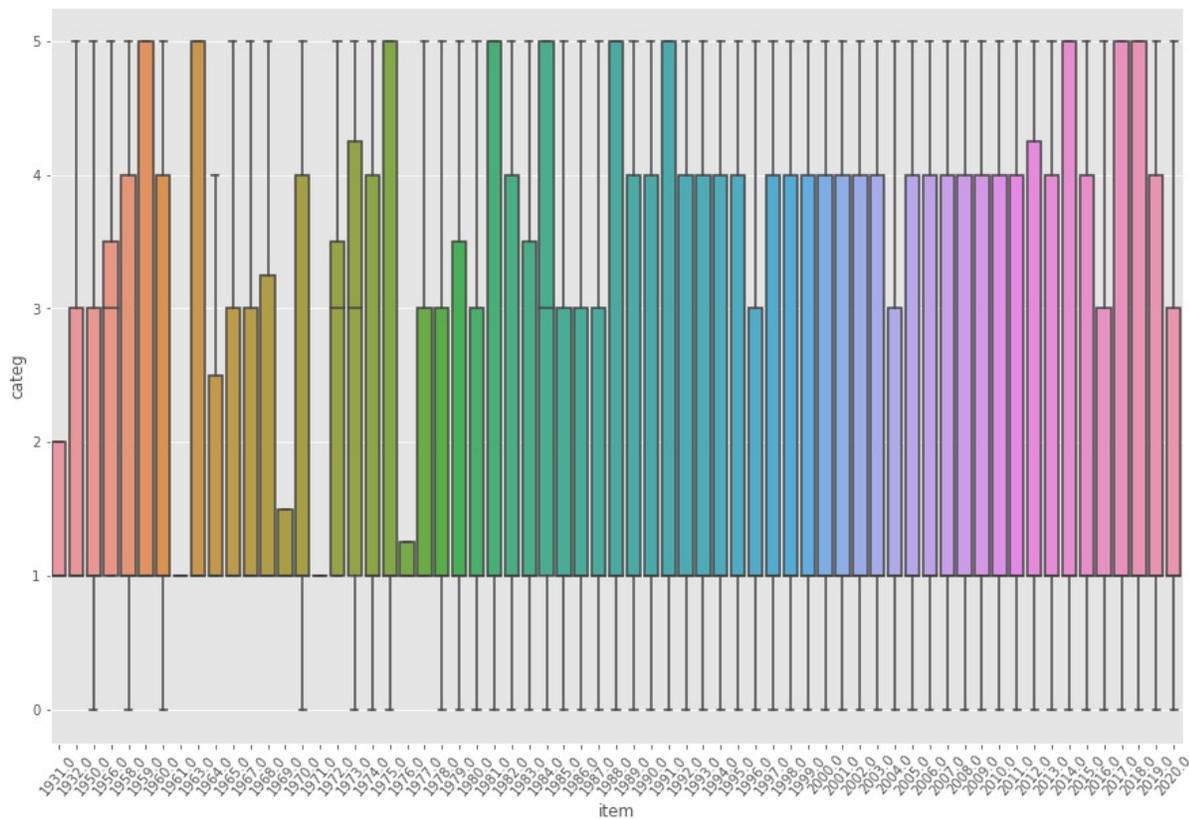


Figura 4-10: Duelo en categoría de Búsqueda: Cajas y Bigotes

Como conclusiones generales, es necesario resaltar la incidencia del duelo especialmente en los años 2000 y 2002. Lo cual podría deberse a la agudización del conflicto armado y su repertorio de violencia entre 1996 y 2002. De la misma manera, es claro que la intensidad del duelo varía entre las categorías de búsqueda y de hallazgo; siendo la categoría de búsqueda la

que presenta palabras con mayor intensidad o bien más palabras dentro de la categoría. Así pues, el duelo para los familiares de los desaparecidos en Colombia se narra en mayor medida a partir de la búsqueda, desde la indagación para dar con el paradero de un ser querido. “Enfrentados a una desaparición, las familias se ven obligadas a modificar el curso normal de un duelo” [6]. El duelo para los familiares de los desaparecidos en Colombia adquiere un tinte de pesquiza, ya que al no tener un cuerpo y mucho menos certeza se debe recurrir a acciones y métodos para gestionar la falta y el vacío que deja un familiar. “Una vez se da comienzo a la búsqueda por anfiteatros, clínicas, cárceles, la calle, se sabe finalmente que la persona a quién han buscado hasta en el último rincón se convirtió en una víctima más de esta guerra sucia” [6].

Así como hay mayor incidencia de intensidad para la categoría de búsqueda dentro del duelo y no para la de hallazgo, resta por decir que no es tan fácil encontrar a un desaparecido en Colombia, que las acciones emprendidas no son tan fructíferas y que así mismo los sentimientos que reposan no son de certeza, ni de paz, ni tranquilidad. Por el contrario, la búsqueda de un desaparecido es una acción inacabada pues no se sabe la duración y mucho menos el resultado de esta. Como reflexión final, los resultados de este trabajo, si bien no son excluyentes y mucho menos definitivos, señalan una alta influencia de familias buscando a sus familiares, que es lo mismo que decir que hay muchos de desaparecidos producto del conflicto armado que siguen enterrados en fosas comunes y miles de hijos, hijas, esposos, esposas, padres y madres luchando por encontrarlos.

Por otro lado, si bien no se habla con la misma intensidad desde la categoría de hallazgo, esta sin duda representa “una luz de esperanza para las familias de las víctimas que yacen allí” [7, p. 367]. Por este motivo, buscar al desaparecido se convierte en “la única posibilidad de tener certeza sobre el paradero de los desaparecidos, indagar por las circunstancias, causas y la identidad de los involucrados en actos atroces, para que tanto las personas cercanas a las víctimas allí enterradas como la sociedad colombiana en general, empiecen a conocer la verdad de lo ocurrido” [7, p. 367].

4.3. Análisis de estigmatización de comunidades étnicas en Buenaventura

Este informe fue culminado el 28 de enero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. El presente documento tiene como interés principal presentar un análisis basado en el corpus compuesto por las entrevistas hechas en Buenaventura que están relacionadas con los grupos étnicos y las dinámicas que estos cumplen en el territorio ya designado. Lo anterior con el interés de profundizar y conocer en mayor medida la forma en la que el conflicto armado ha afectado a estas comunidades, en especial las comunidades negras, en

un territorio tan disputado y relevante en Colombia como lo es Buenaventura. Es así como a través de este análisis se busca observar e identificar de qué manera se presenta o no la estigmatización de esta comunidad y las razones por las cuales esta se da. Esto con el objetivo de presentar un manejo y análisis de datos que haga posible la determinación de bases para futuros procesos de investigación, el esclarecimiento y entendimiento de lo sucedido en el marco del conflicto armado en Colombia.

A partir del núcleo 7 del árbol de entrevistas conocido como Causas, dinámicas e impactos del conflicto armado sobre los pueblos y territorios étnicos, se filtraron las siguientes subcategorías: Racismo y discriminación étnico-racial; Relación entre el Estado y los Pueblos Étnicos; Procesos de colonización en territorios étnicos; Relaciones interétnicas o interculturales; Tensiones identitarias. Estas subcategorías seleccionadas comprenden las relaciones entre los distintos actores y qué tipo de dinámicas se dan entre ellos para evidenciar si hay algún tipo de estigmatización o forma de discriminación presentes y también encontrar cómo en acciones como la colonización de territorios étnicos si se está repitiendo algún tipo de práctica que evidencie la existencia de discriminación ya sea internalizada o estructural que esté marcando y legitimando la estigmatización de estas comunidades. A través del procesamiento del corpus, se encontró un resultado promedio de 247.5 tokens aproximadamente.

Para complementar el análisis se construyó un lexicón, el cual consiste en una lista categorizada del vocabulario referente al tema a tratar, es decir que es un tipo de diccionario que cuenta con los términos más relevantes que son clasificados en categorías definidas para establecer una estructura semántica lo más amplia posible para poder contrastarla con la información contenida en el corpus, para poder así, encontrar relaciones relevantes que aporten al análisis. Es así como el lexicón generado cuenta con términos relacionados con las comunidades étnicas, las formas de discriminación, los diferentes términos que se relacionan con la raza y la experiencia del racismo. De igual forma, dentro de este lexicón se encuentran adjetivos o expresiones que pueden referirse a situaciones en las que se está presentando una descripción de connotación estigmatizadora o discriminatoria.

Algunos de los elementos que se tuvieron en cuenta para construir el lexicón consistieron en otros lexicones o listas de palabras tomadas de la Universidad de Toledo y el Algonquin College de Ottawa. De igual forma se tomaron como referente textos como *Pieles negras, máscaras blancas* (1952) de Frantz Fanon, *Identidades a flor de piel* (2003) de Elisabeth Cunin, *Talks and disputes of racism in Colombia after multiculturalism* (2018) y *Etnización de la negritud: la invención de las “comunidades negras” como grupo étnico en Colombia* (2013) ambas obras de Eduardo Restrepo en donde se habla directamente de las comunidades negras, su conformación y los discursos que se dan en torno a su identidad y las complejidades que se establecen en torno a su performatividad en el país.

Una vez tenido el lexicón establecido, para realizar un análisis más completo del corpus se hizo una categorización de los términos contenidos en el lexicón generado. Lo anterior para poder así conocer a forma de niveles y bajo categorías determinadas, la manera en la que en las entrevistas se desarrolla el tema de la discriminación. A continuación, se muestran las categorías usadas:

- Categoría de estigmatización verbal: Dentro de esta categoría se organizan los términos que se pueden encontrar en las entrevistas que muestren que aquel que está hablando se encuentra de alguna manera reforzando y mostrando tener un discurso discriminatorio y estigmatizador.
- Categoría de estigmatización tribal: Esta categoría se refiere a los términos del lexicón cuya connotación se relacione con la estigmatización o discriminación. Se encuentra determinada por la pertenencia a un grupo específico, que en este caso serían los grupos étnicos, y las características específicas del mismo.
- Categoría de estigmatización estructural: Los términos categorizados como estigmatización estructural son aquellos términos que presentan la idea de que estas formas de discriminación se encuentran presentes en la construcción social, política y cultural del país, por lo cual estas palabras que contienen la idea estigmatizadora se encuentran presentes en la forma en la que se habla de leyes, o en la forma en la que se organizan instituciones, ya sea privadas o del Estado.
- Categoría no directamente estigmatizadora: Los términos de esta categoría son aquellos que son relevantes para la comprensión de la estigmatización y la discriminación pero que en sí no tienen una connotación de este tipo.

La posterior clasificación por grados de intensidad es la siguiente:

- 0: Nada relevante: Palabras que no aplican para una categoría específica.
- 1: Muy poco relevante: Palabras que por sí solas no tienen relación evidente con la categoría. Sin embargo, resultan útiles dentro de los análisis.
- 2: Poco relevante: Palabras que, aunque tienen que ver con la categoría no son específicas del tema y pueden corresponder a otros similares. Sin embargo, resultan útiles para analizar de lo general a lo específico.
- 3: Medianamente relevante: Palabras que denotan algún tipo de estigmatización de forma leve. Están asociadas a las palabras alusivas a la discriminación.
- 4: Relevante: Palabras asociadas fuertemente con la discriminación o la estigmatización. Se relacionan de forma evidente con la estigmatización.

- 5: Muy relevante: Palabras específicamente relacionadas con la estigmatización especificada en la categoría. En especial aquellas que buscan responder a la pregunta de investigación.

Para la presentación de los resultados, primero se exponen las frecuencias de algunos términos en una gráfica de barras acompañada por una nube de palabras. Un ejemplo de esto se encuentra en la imagen 4-11 donde se presentan los verbos más frecuentes dentro del corpus, los cuales son decir, tener, hacer, ir, llegar, entre otros. Esta información es representada de manera diferente en la imagen 4-12, dónde se presenta la nube de palabras de todos estos verbos en diferentes colores. Entre mayor sea el tamaño de las palabras, mayor presencia tiene dentro del corpus.

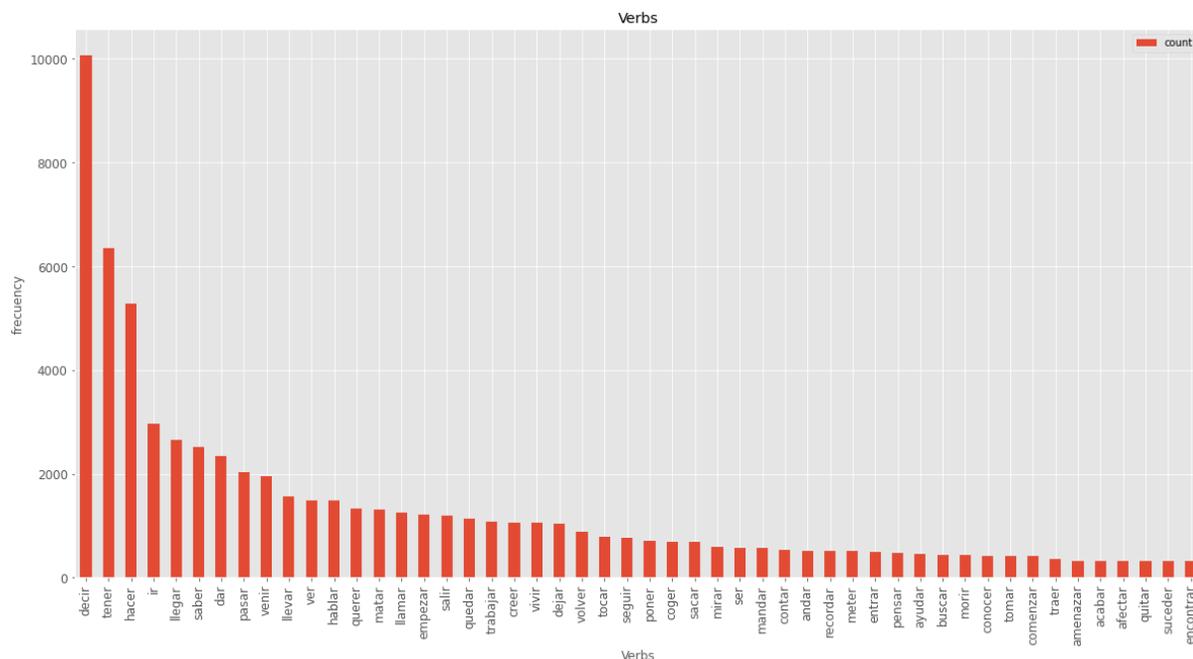


Figura 4-11: Gráfica de los 50 verbos más frecuentes



Figura 4-13: Gráfica modelo Fast Text

Por último, para la presentación de resultados de las categorías desarrolladas para este informe en específico, se obtiene la imagen 4-14, la cual muestra el análisis del valor de las intensidades de los términos del lexicon por año en relación con la categoría de estigmatización verbal. El año en el que hay mayor intensidad corresponde al 2004, la intensidad de la estigmatización verbal en este caso sobrepasa el valor de 1600, los siguientes años con valores más altos fueron el 2002, el 2016 y el 2005. Entre el 2000 y el 2004, fue una época relevante en Buenaventura ya que es cuando comienzan a llegar los paramilitares a la zona con el objetivo de adquirir control y dominar las rutas de narcotráfico de la zona por lo que esta época se caracterizó por la victimización de la población y la perpetración de masacres

[8]. El 2005 fue el año posterior a la desmovilización del Bloque Calima, que fue reconocida como una desmovilización fallida ya que muchas personas “se rearmaron, iniciaron cruentas disputas entre sí...” [8] esto con el fin de establecer nuevos mandos y controlar los negocios ilegales. El año 2016 cuenta con un valor de intensidad de más de 1200, cosa que puede estar relacionada con la forma en la que se llevó a cabo el Proceso de Paz en el puerto.

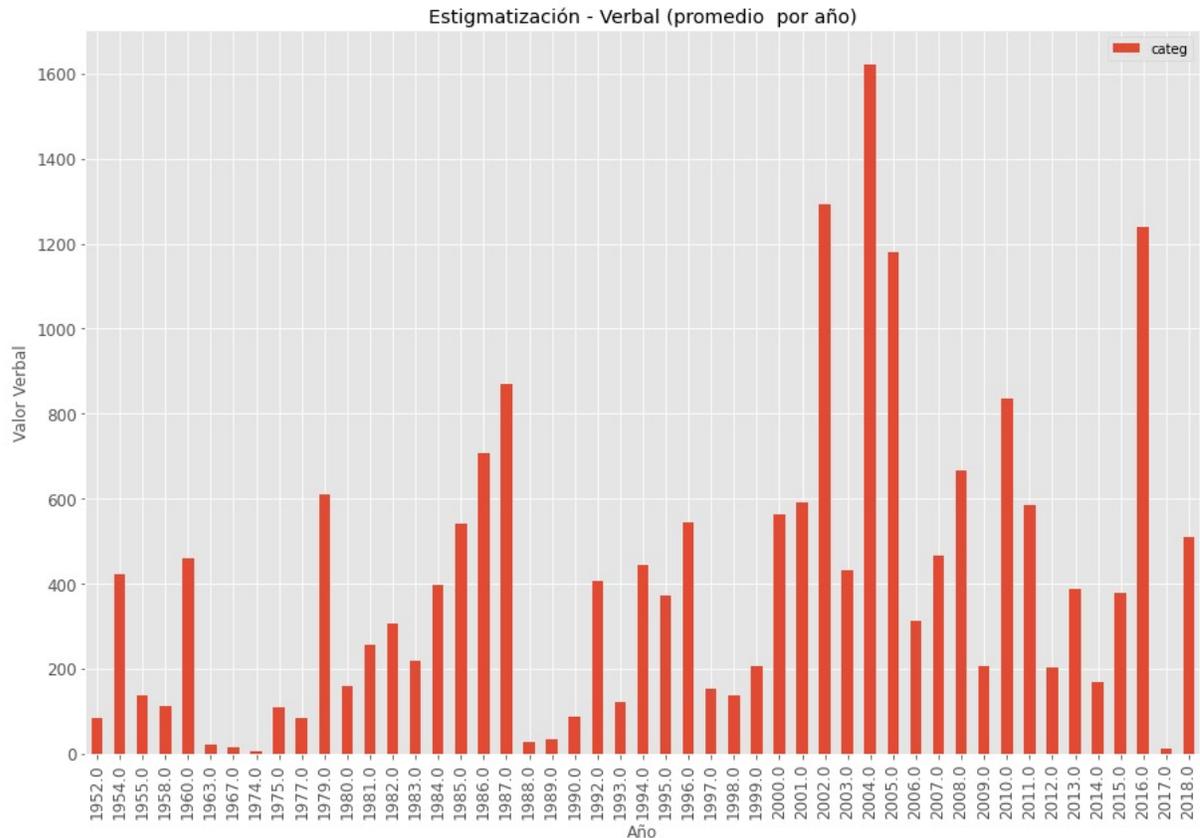


Figura 4-14: Gráfica de análisis histórico del promedio de frecuencia de términos relacionados con la estigmatización verbal

Otra forma de la representación de los datos de esta categoría en específico es a través de la gráfica de caja de bigotes vista en la imagen 4-15, dónde se presenta el promedio de la intensidad de los términos relacionados con la estigmatización verbal. Se puede evidenciar que los términos, en su gran mayoría, se encuentran en niveles de intensidad entre 3 y 5, lo cual evidencia que en la temporalidad tenida en cuenta se establecen términos en las entrevistas relacionados con niveles medios y altos de estigmatización verbal. Los años comprendidos dentro de 1954, 1955, 1958, 1990, 2007 y 2008 cuentan con términos de todas las intensidades. En cambio, en años de 1984 a 1987 y después en 1989 la mayoría de los términos de esta categoría tenían una intensidad de entre 4 y 5, también en 1993, 1994, 1996, 1997 del 2000 al 2002, 2004 y 2006, del 2008 al 2013 y el 2015. Esto es algo que vale la pena analizar para comprender por qué se presenta este nivel de estigmatización verbal en estos años.

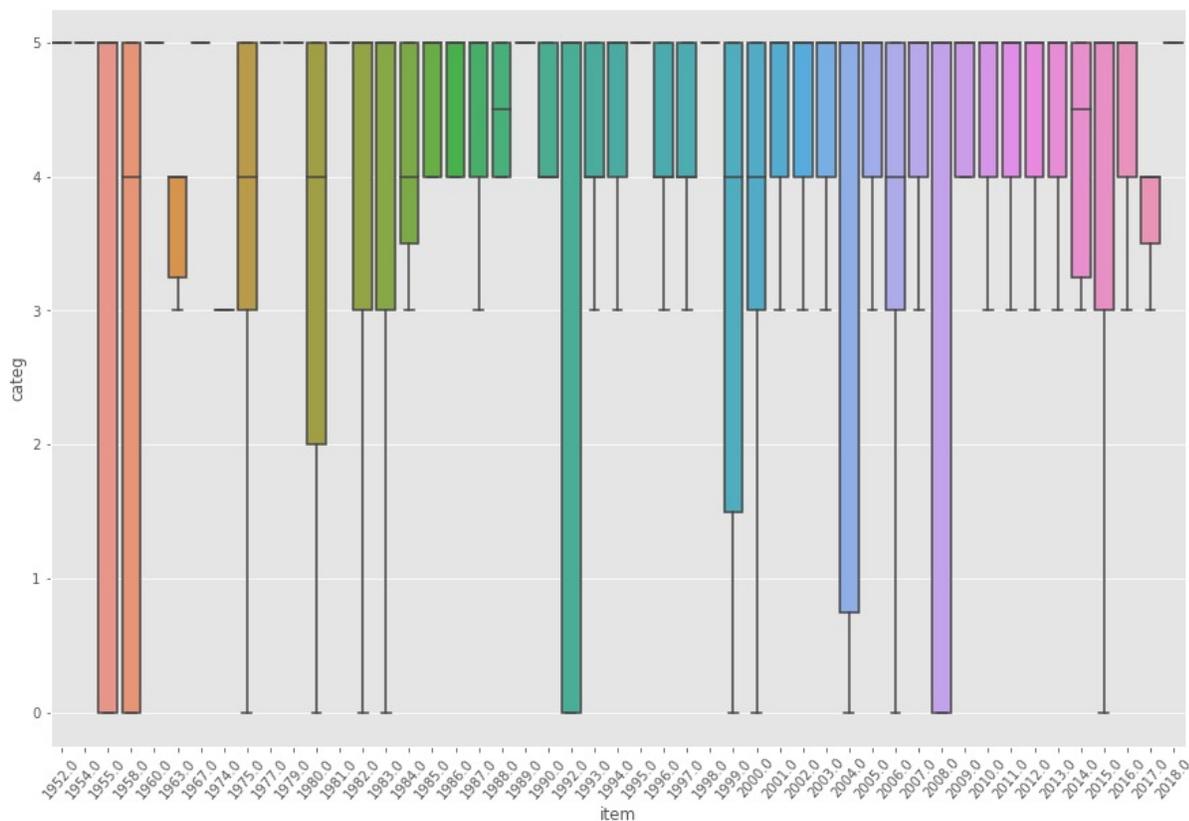


Figura 4-15: Gráfica de promedio de intensidad de los términos de la muestra

Como conclusiones generales, se puede evidenciar que a través de los diferentes resultados obtenidos se puede realizar un análisis más profundo de la forma en la que la estigmatización se da en la zona de Buenaventura. En primer lugar, el análisis de frecuencias puede ofrecer un acercamiento a los temas que se tratan de forma general en las entrevistas, permitiendo cuestionarse sobre las distintas connotaciones que tiene cada una de estas palabras. En el caso de los adjetivos, verbos y chunks frecuentes analizados, se pudieron encontrar términos que pueden ser relevantes para el análisis de la estigmatización, ya que en el caso de los adjetivos se encuentra el término bueno que puede mostrar una dicotomía entre aquello socialmente considerado bueno. En el caso de los chunks, se encuentran términos que desarrollan la noción de grupo como gente, comunidad y cabildo, esto es relevante debido a que la estigmatización se genera por un atributo diferencial que en muchos casos se generaliza a grupos específicos, razón por la cual analizar estas nociones resulta importante para comprender de qué forma se habla sobre estas. Finalmente, en lo que respecta a verbos, estos resultan un poco más complejos de analizar, pero, considerando el tema tratado, verbos como matar y afectar pueden relacionarse con los efectos directos del conflicto, pero si se profundiza, se puede analizar sobre los receptores de estas acciones para poder identificar si se trata de estigmatización.

4.3 Análisis de estigmatización de comunidades étnicas en Buenaventura

En relación con el análisis hecho con el uso del lexicón, se pueden establecer relaciones específicas de las formas de estigmatización con las temporalidades a las que se refieren las entrevistas del corpus. Antes de profundizar en los momentos específicos más relevantes de los resultados, es de relevancia comprender de forma general los años que estuvieron comprendidos en las gráficas ya que estos años van determinados desde 1952 hasta el 2018. En aspectos generales, se puede inferir que dentro de las temporalidades se presentan épocas violentas de la historia del país. Considerando que el primer año tomado en cuenta es 1952, año donde se continúa perpetuando la conocida Violencia. Según la Biblioteca Nacional de Colombia se establece que esta época (1946-1958) se caracterizó por ser la suma de violencias políticas, económicas, sociales y religiosas [9].

Posterior a esta época, se instaura el Frente Nacional. En los 60, las tierras del pacífico eran reconocidas como tierras baldías o reservas, en las cuales, en realidad se encontraban comunidades de campesinos negros que, aunque se organizaron en comunidades, el Estado aún no reconocía sus derechos a la tierra por lo que en esta década comenzaron a otorgar tierras con el interés de permitir a sujetos privados explotar “bajo un modo capitalista y moderno” [10, p. 148] las tierras. Ante esto, las décadas de los 60 y los 70 en el pacífico se caracterizaron por la discusión colectiva sobre el territorio y la manera por medio de la cual exigir por la misma ya fuera de forma comunitaria o individual. en los 70 se agudiza más la discusión en relación a las reivindicaciones identitarias y por ende las movilizaciones de carácter étnico [10, p. 214].

Los 80 se caracterizan por ser la época en la que comenzaron a llegar los grupos armados a las diferentes zonas del Pacífico. En el caso de Buenaventura, las FARC llegaron a la zona en donde la presencia de estos se convirtió en algo normalizado [8, p. 157], ya que estos, al comprender sobre las dinámicas de narcotráfico de la zona, decidieron adoptar esto como parte de su financiación. Esta dinámica se mantuvo hasta el año 2000, que fue el momento en el que los paramilitares llegaron a Buenaventura haciendo que los próximos 4 años se caracterizaran por actos de sevicia y constantes masacres por las diferentes disputas por el control del territorio y las rutas de narcotráfico. En el año 2004, que es de alta relevancia en los resultados del análisis del corpus, se firmó un acuerdo de desmovilización del Bloque paramilitar Calima, este acuerdo fue cuestionado porque, aunque no se escuchaba del Frente Pacífico, la presencia armada que dejó este se mantuvo [8, p. 186].

“Los años de 2005 a 2007 fueron recordados como años de locura...” [8, p. 187]. Durante estos años, la violencia de la zona se caracterizaba por los enfrentamientos entre las FARC y los diferentes grupos rearmados como los Rastrojos y otros grupos creados como los Urabeños. Estos años fueron aquellos en los que la violencia en el territorio fue más alta, hasta el 2010 que se crea el grupo armado ilegal La Empresa. Es así como la presencia de todos estos actores hacen que Buenaventura se convierta en un territorio violento y de alta complejidad, ya que hay enfrentamientos entre bandas, grupos armados y la fuerza pública, generando así

altos niveles de victimización. Es así como para el periodo comprendido entre 1990 y 2014, 152.837 personas fueron víctimas del desplazamiento forzado y en Buenaventura cerca de 1632277 personas fueron víctimas de violencia según la RNI [8, p. 213]. Finalmente, otro año relevante para este año fue la firma del acuerdo de paz en el año 2016.

Bibliografía

- [1] Oxford, “Oxford languages,” 2021.
- [2] A. Moreno Ortiz, “El lexicon en la lexicografía computacional: adquisición y representación de información lexica,” 2000.
- [3] fastText, “fasttext:library for efficient text classification and representation learning,” 2020.
- [4] G. Grupo de Memoria Histórica, “¡basta ya! colombia: Memorias de guerra y dignidad,” 2013.
- [5] T. D. V. Catalogue’, “Caja y bigotes,” 2021.
- [6] C. G. de Piñeres, “Procesos de duelo en familiares de víctimas de desaparición forzada,” vol. <https://psicologiajuridica.org/psj137.html>, s.f.
- [7] S. y. P. M. Carrero, “Hallazgo de fosas comunes en colombia. el tiempo de las víctimas,” vol. Revista Criminalidad, Vol 50, Número 1, ISSN 1974-1974, 2008.
- [8] C. N. de Memoria Histórica, “Buenaventura: Un puerto sin comunidad,” 2015.
- [9] A. Caballero, “Historia de colombia y sus oligarquías (1498 - 2017),” s.f.
- [10] O. Hoffmann, “Comuniadades negras en el pacífico colombiano: innovaciones y dinámicas étnicas,” 2007.

5 Lexicones

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las entrevistas de los diferentes corpus utilizados dentro de los diversos informes que comprenden múltiples objetivos y preguntas de investigación.

El trabajo metodológico se entiende por medio del procesamiento de entrevistas pertenecientes al corpus recolectado por la Comisión de la Verdad bajo su mandato de acuerdo con los diversos intereses de los diferentes informes. El proceso se desarrolla con ayuda de lexicones creados para los temas objetivos que ayuden a clasificar los diferentes aspectos por investigar con valores cuantitativos. Es importante mencionar que los datos transformados pueden vincular aspectos cualitativos, permitiendo investigar desde lo positivista y científico para ocasionar resultados de índole cualitativa.

Los resultados se presentan de tres maneras diferentes:

- Gráficas de barras por categorías: Este tipo de resultado presenta las frecuencias de las diferentes categorías analizadas, junto con las posibles subcategorías provenientes de la categoría que permitan visibilizar las diferentes dimensiones a las que se somete el análisis.
- Tablas por categorías: Este tipo de resultado presenta los datos que también se encuentran en las gráficas de barras por categorías. La diferencia se encuentra en la presentación de la información, puesto que puede ser más exacta la información de las tablas al presentar el número exacto del valor suministrado.
- Gráficas de intensidad: Este tipo de resultado compara de a dos categorías y las entrevistas que dan cuenta de estos, para encontrar tanto similitudes como diferencias en las intensidades según los eventos a los que estuvo expuesto el entrevistado.

5.1. Análisis del lenguaje afectivo utilizado en entrevistas

Este informe fue culminado el 1 de febrero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. El objetivo de este texto es analizar el lenguaje afectivo usado en diferentes entrevistas realizadas por La Comisión de la Verdad, con el fin de encontrar patrones y tendencias del lenguaje utilizadas por aquellos que narran sus experiencias de vida en torno

al conflicto armado colombiano, y cómo este uso del lenguaje se relaciona específicamente con los diferentes hechos victimizantes dentro del marco de la guerra interna colombiana. Para este propósito, a través de técnicas de procesamiento del lenguaje, se identificarán las categorías emocionales más frecuentes a la hora de narrar hechos victimizantes, así como niveles de intensidad dentro de los relatos agrupados por hecho; utilizando una construcción de lexicones relacionados con el lenguaje afectivo.

Los lexicones utilizados fueron dos. El primero proporciona distintas categorías emocionales y aquellas palabras que están relacionadas con estas categorías. Para esto se utilizó, por un lado, el lexicon SEL (Spanish Emotion Lexicon), que cuenta con 2036 palabras, asociadas a las emociones básicas: alegría, enojo, miedo, tristeza, sorpresa, y repulsión. Para lograr esta clasificación, se necesitaron 19 anotadores y cada uno evaluaba que tan probable era la asociación de determinada palabra con cada una de las emociones recientemente descritas. Posteriormente, se promediaron los valores proporcionados por los anotadores, para luego clasificar cada palabra dentro de una emoción. Por otro lado, para la construcción del lexicon usado para este trabajo, se realizó la traducción del lexicon NCR, donde los términos se asocian a 8 diferentes emociones: enojo, miedo, alegría, sorpresa, tristeza, disgusto, confianza (como categoría macro, por lo que son entendidas tanto positivas como negativas), y expectativa. Al igual que con el SEL, para el NCR, cada palabra fue evaluada por 5 personas diferentes y así ubicar los términos dentro de las emociones correspondientes. Consecuentemente, se trabajó con la compilación de estos dos lexicones, con un total de 6.626 palabras.

El segundo es una extensión del recurso expuesto anteriormente. Además de proporcionar las categorías cualitativas, propone un análisis cuantitativo. Para su realización se compilaron y posteriormente tradujeron al español, distintos recursos afectivos encontrados en inglés: NCR, General Inquirer, ANEW, XANEW. El NCR, además de lo explicado en el lexicon anterior, le otorga a cada palabra un valor de intensidad, promediando las respuestas de los observadores correspondientes, que va de 0 a 1 (siendo 1 el nivel de mayor intensidad). El General Inquirer asigna a cada palabra el POS (part of speech) y asocia cada término con polaridad positiva o negativa según corresponda. Por otro lado, el ANEW y XANEW indica para cada palabra valores, de 1 a 9 (siendo 1 el valor más bajo) de “Valence”, “Arousal” y “Dominance”, donde el primero hace referencia a lo placentero, el segundo a la intensidad de la emoción y el último al grado de control ejercido por un estímulo. Esto se logró promediando el valor que cada participante del trabajo inicial le dio a cada término para estas tres diferentes categorías numéricas. Finalmente, se utilizó, nuevamente, el recurso SEL usado en el primer lexicon, y sus valores numéricos de factor de probabilidad de uso afectivo para cada palabra, de 0 a 1 (siendo 0 el nivel más bajo de intensidad).

A continuación, se presentarán varios ejemplos de los resultados después del procesamiento con lenguaje natural. Estos corresponderán a las categorías de Tortura y Otros Tratos Crue-

les, Detención Arbitraria, Homicidio, Desplazamiento Forzado, Ataque a bien protegido y Desaparición Forzada. Para el primer ejemplo, se encuentra a disposición la imagen 5-1 y la tabla 5-1 las cuales muestran la tortura o otros tratos crueles, que se refiere a un sujeto: “que inflija a una persona dolores o sufrimientos físicos o psíquicos, con el fin de obtener de ella o de un tercero información o confesión, de castigarla por un acto por ella cometido o que se sospeche que ha cometido o de intimidarla o coaccionarla por cualquier razón que comporte algún tipo de discriminación” [1]. Después de procesar los textos respectivos para este hecho victimizante, se puede notar un patrón similar entre los hechos correspondientes ilustrados en esta gráfica, que las mayores frecuencias ocurren en sorpresa (1074), tristeza (774) y, posteriormente, el enojo (473). Las respuestas emocionales menos representadas fueron el disgusto, la confianza y la expectativa.

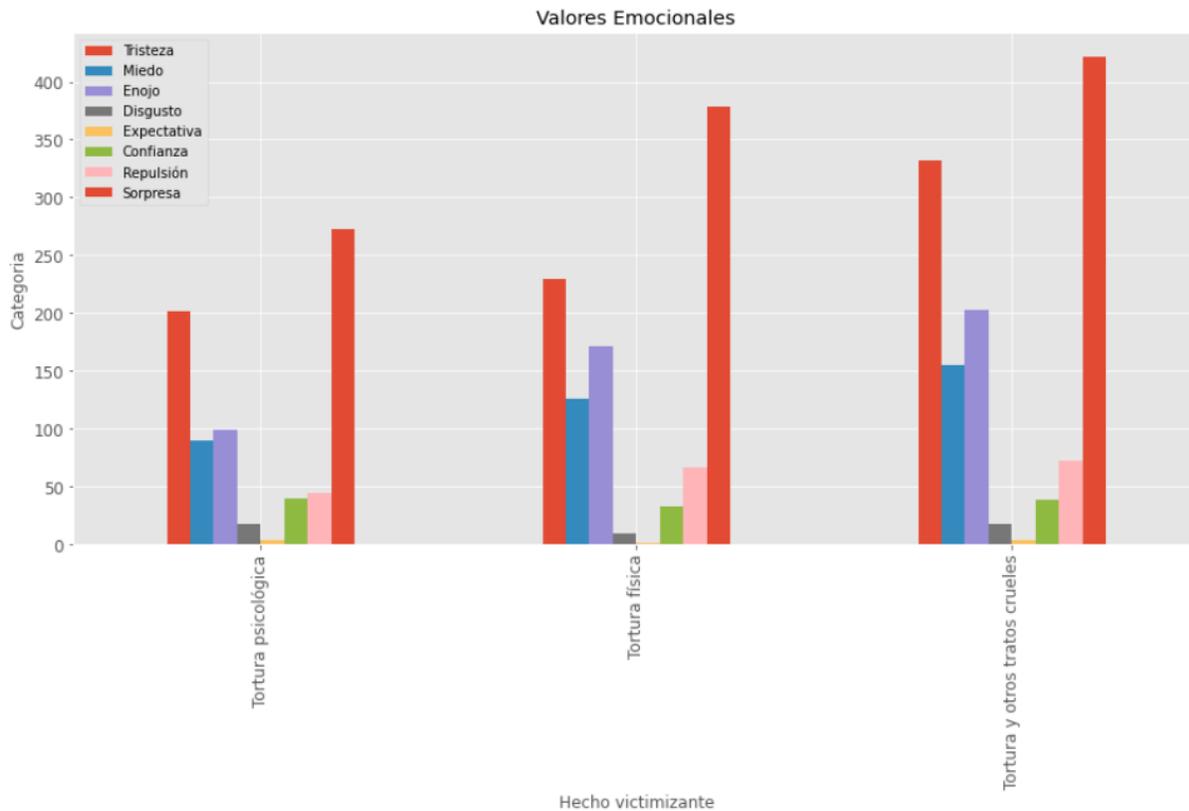


Figura 5-1: Categorías emocionales en “Tortura y otros tratos crueles”

Tabla 5-1: Categorías emocionales en “Tortura y otros tratos crueles”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Tortura psicológica	99	4	18	273	40	44	90	202
Tortura física	171	1	9	379	33	67	126	230
Tortura y otros tratos crueles	203	3	18	422	39	72	155	332

Como particularidades de este patrón general, se puede observar que la frecuencia relativa proporcionalmente del enojo es un poco mayor que cuando se trata de tortura física. Por su parte, en la tortura psicológica, el miedo está más cerca del enojo. Probablemente, las características mismas del tipo de tortura expliquen esta diferencia, porque la vulneración directa de la integridad física podría generar más rabia que el temor a lo desconocido implícito en la tortura psicológica, por su carácter más mediato y menos predecible.

Para el segundo ejemplo se encuentran a disposición la imagen **5-2** y la tabla **5-2** que se refieren a cuando un “servidor público que sin el cumplimiento de los requisitos legales reciba a una persona para privarla de libertad o mantenerla bajo medida de seguridad” [2]. Dentro del análisis, demuestran la notoria tendencia hacia el uso de palabras categorizadas como pertenecientes a la sorpresa (41 % de los términos encontrados en el lexicón pertenecen a esta categoría) y, seguidamente, de la tristeza. En el caso de miedo y enojo, son muy parejas las frecuencias de uso. Ahora bien, en el caso de detención tras cumplimiento de pena, comparativamente, las frecuencias cambian, teniendo en primer lugar el enojo, el disgusto y la sorpresa. Esto se podría dar, porque después de cumplir con una condena, lo último que se espera que es vuelvan a detener a la persona en cuestión por la misma acción delictiva, lo que posiblemente, sería a causa de la respuesta de emociones después de terminar de cumplir con una pena, lo que se materializa en el enojo y la sorpresa. Esto mismo podría servir para hacer hipótesis de la baja frecuencia del miedo en este hecho victimizante, pues tener una experiencia previa puede amortiguar el temor, al haber menos incertidumbre.

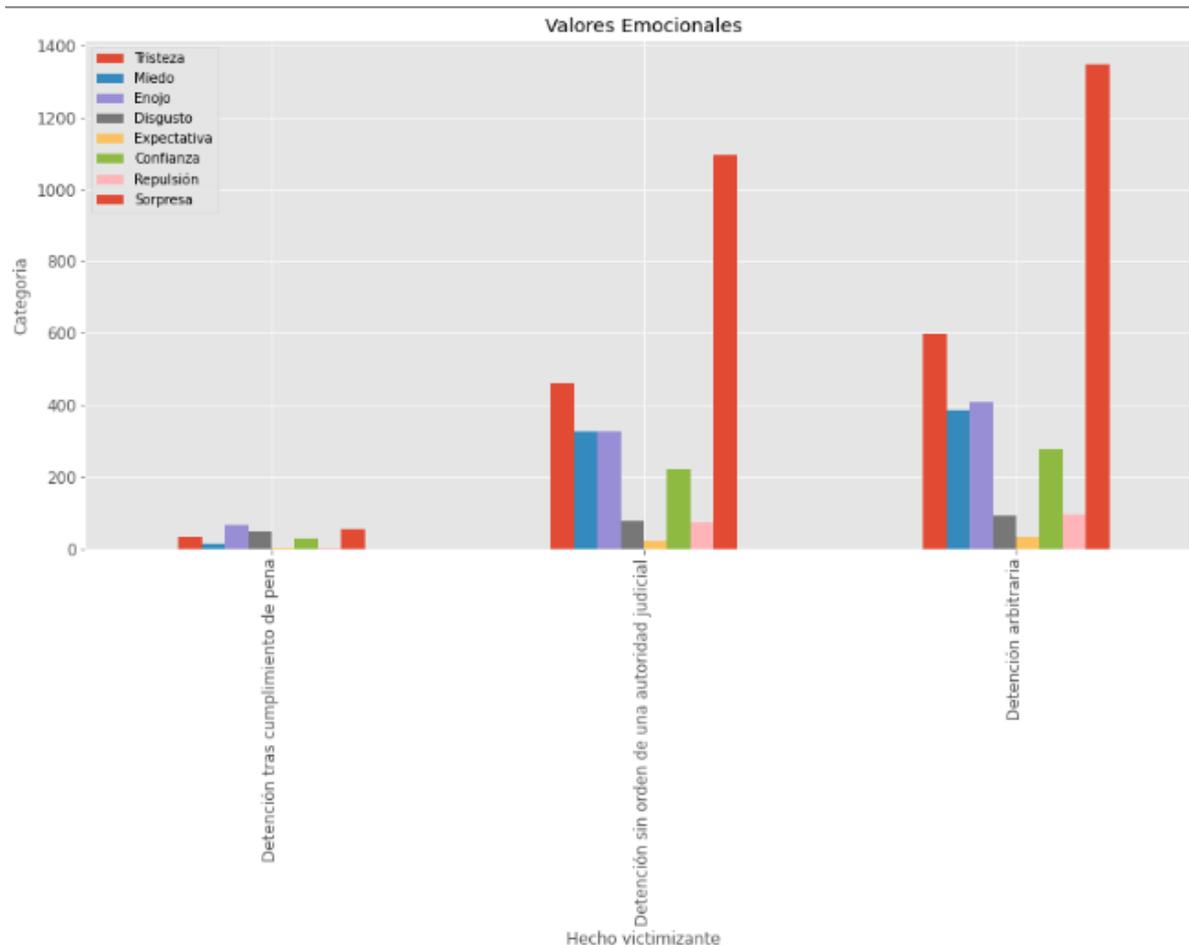


Figura 5-2: Categorías emocionales en “Detención arbitraria”

Tabla 5-2: Categorías emocionales en “Detención arbitraria”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Detención tras cumplimiento de pena	64	1	46	54	30	1	15	33
Detención sin orden de una autoridad judicial	327	20	77	1097	220	74	327	460
Detención arbitraria	407	33	92	1347	279	94	384	596

Para el tercer ejemplo se encuentran a disposición la imagen **5-3** y la tabla **5-3**, las cuales demuestran la categoría emocional de homicidio. Después de procesar las entrevistas relacionadas con homicidio como macro categoría, es evidente que la tristeza y la sorpresa son las respuestas emocionales más frecuentes (aproximadamente 28 % y 30 % respectivamente). Es importante tener en cuenta que cuando se habla de homicidio, la víctima es forzosamente indirecta, un familiar, amigo o allegado de la víctima directa. El que esta conducta violenta haya afectado a una persona cercana, la imposibilidad de enfrentar o confrontar al sistema

violento que ha ejecutado la conducta, explica el predominio de la tristeza. Siguiendo con el caso de la sorpresa, esto se podría explicar porque al ser una muerte violenta no es esperada por sus allegados.

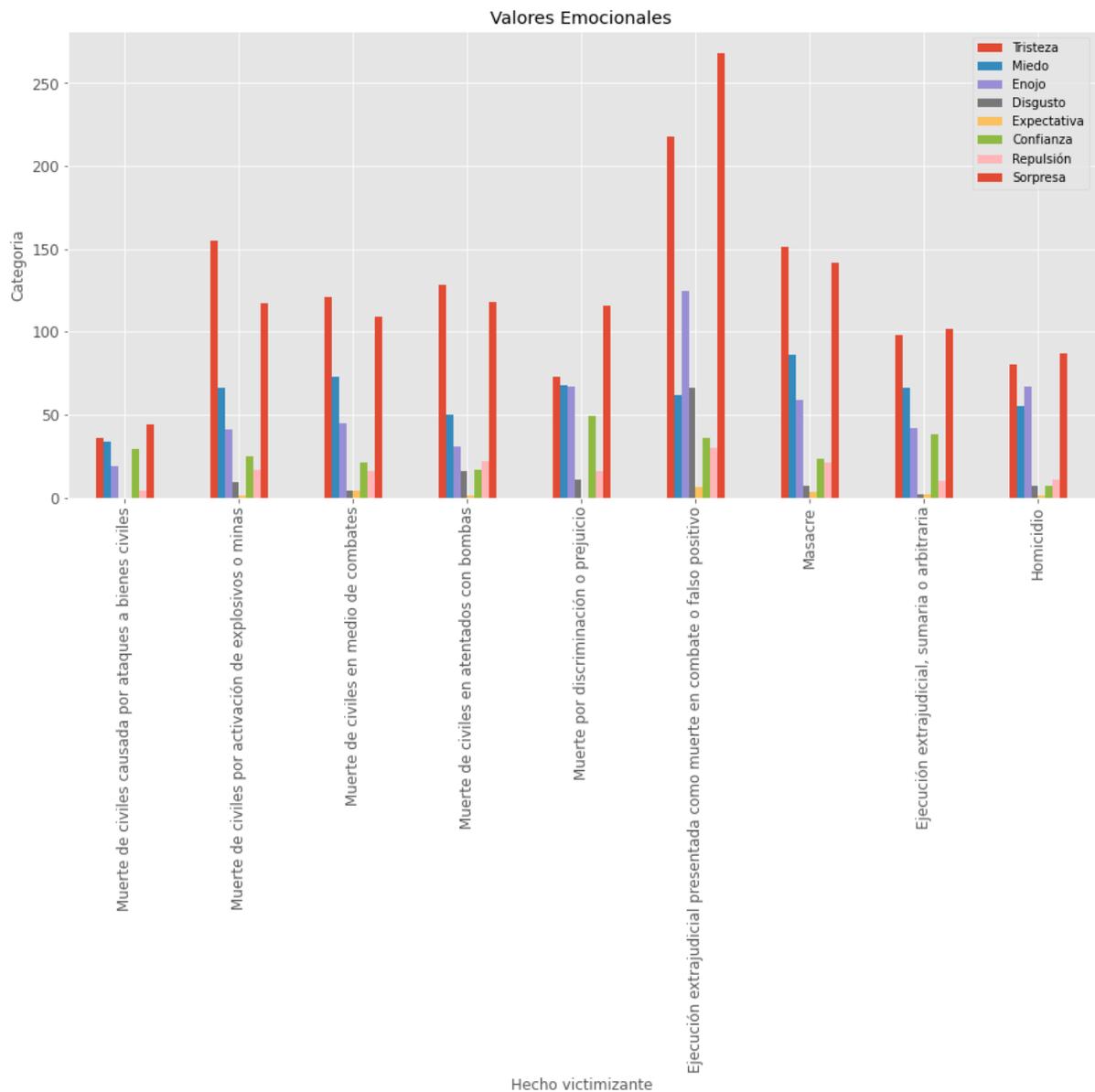


Figura 5-3: Categorías emocionales en “Homicidio”

Tabla 5-3: Categorías emocionales en “Homicidio”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Muerte de civiles causada por ataques a bienes civiles	19	0	0	44	29	4	34	36
Muerte de civiles por activación de explosivos o minas	41	1	9	117	25	17	66	155
Muerte de civiles en medio de combates	45	4	4	109	21	16	73	121
Muerte de civiles en atentados con bombas	31	1	16	118	17	22	50	128
Muerte por discriminación o prejuicio	67	0	11	116	49	16	68	73
Ejecución extrajudicial presentada como muerte en combate o falso positivo	125	6	66	268	36	30	62	218
Masacre	59	3	7	142	23	21	86	151
Ejecución extrajudicial, sumaria o arbitraria	42	2	2	102	38	10	66	98
Homicidio	67	1	7	87	7	11	55	80

Es interesante el caso de la ejecución extrajudicial presentada como muerte en combate o falsos positivos, pues la frecuencia de la sorpresa y la tristeza es alta, así como el enojo. Esto se podría explicar porque en los casos de falsos positivos, las personas y sus familias fueron engañadas o las víctimas directas fueron seleccionados dentro de grupos vulnerables, que no se esperan que sean blancos de violencia. Si a esto se suma la historia del uso frecuente de este tipo de ataque a lo largo del conflicto armado colombiano (La JEP estima al menos 6402 casos solamente entre 2002 y 2008)[3], se puede agregar un efecto de la memoria que amplifica la respuesta emocional de la tristeza y el enojo.

Para el cuarto ejemplo se encuentra a disposición la imagen 5-4 y la tabla 5-4 que muestran la categoría emocional de desplazamiento forzado. Es decir, cuando “de manera arbitraria, mediante violencia u otros actos coactivos dirigidos contra un sector de la población, ocasione que uno o varios de sus miembros cambie el lugar de su residencia” [4] - la imagen y la tabla arrojan altos niveles de sorpresa y tristeza, seguido del miedo, el enojo y la confianza. Sin embargo, la tristeza es un poco más frecuente cuando se trata de desplazamiento masivo (241 palabras encontradas en el lexicón para esta categoría).

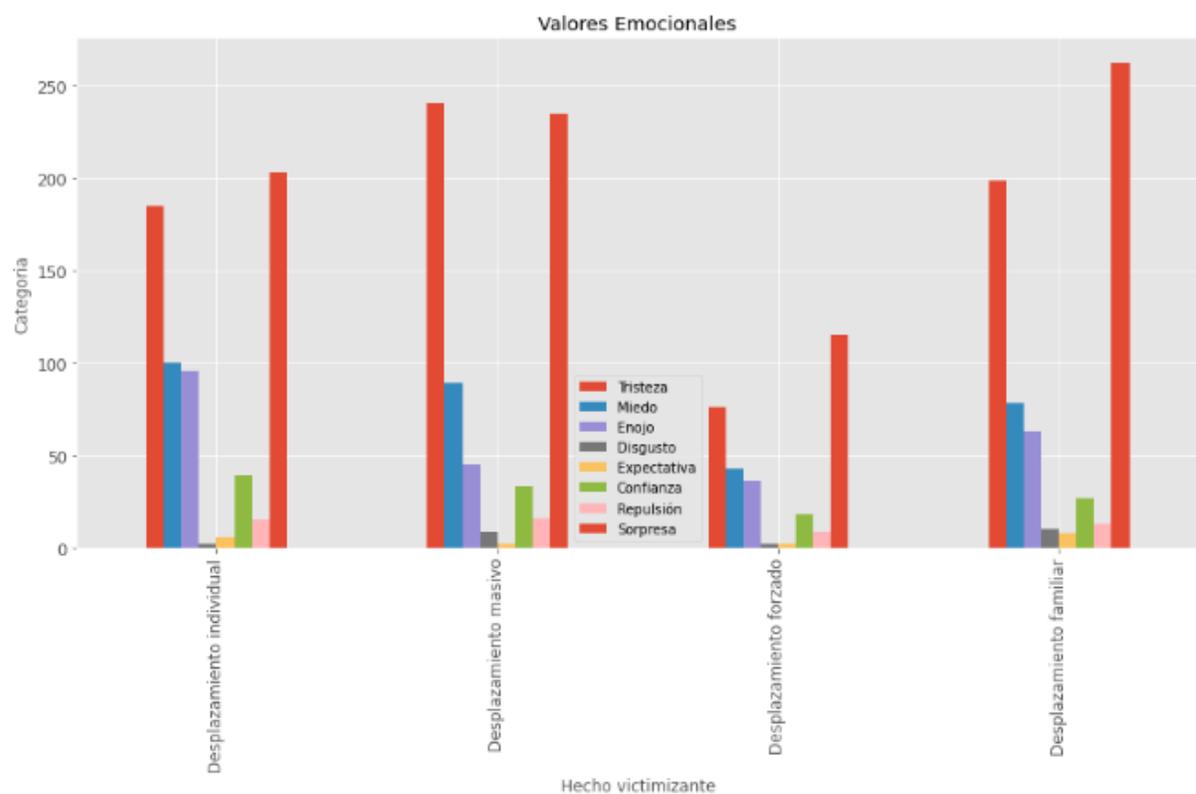


Figura 5-4: Categorías emocionales en “Desplazamiento forzado”

Tabla 5-4: Categorías emocionales en “Desplazamiento forzado”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Desplazamiento individual	96	6	2	203	39	15	100	185
Desplazamiento masivo	45	2	9	235	33	16	89	241
Desplazamiento forzado	36	2	2	115	18	9	43	76
Desplazamiento familiar	63	8	10	263	27	13	78	199

Estos fenómenos se podrían explicar porque no es predecible un desplazamiento, ocurre de manera repentina y quienes experimentaron estos hechos debieron salir con lo que pudieran. En cuanto a la tristeza como la respuesta emocional más común en el desplazamiento masivo, se podría hipotetizar, que no solo se está dejando la propiedad, sino también a todas las personas, costumbres y dinámicas sociales que se viven en una zona, o en muchos casos estas zonas quedaron completamente destruidas. Por otro lado, la tabla muestra que los niveles de miedo y enojo son más altos en desplazamiento individual, comparativamente, con los otros subtipos de este hecho; lo que es posible que suceda porque no se cuenta con una red de apoyo como cuando se trata de un desplazamiento familiar o colectivo.

Para el quinto ejemplo se encuentran a disposición la imagen 5-5 y la tabla 5-5. A partir de los textos analizados, es evidente la presencia de sorpresa, tristeza, miedo y enojo, en

ese orden (918, 660, 386 y 375, respectivamente), como respuestas emocionales frecuentes cuando se trata de un ataque a un bien protegido. Como estos hechos, según la Comisión Intersecretarial de Derecho Internacional Humanitario [5], se presentan en lugares de sanidad, religiosos, culturales u obras, no se espera que ocurra un evento de esta magnitud, lo que explicaría la presencia de la sorpresa, e incluso de la afectación a la confianza (no muy común según los hechos analizados), al ser blanco de ataque un lugar que brinda tranquilidad a quienes asisten.

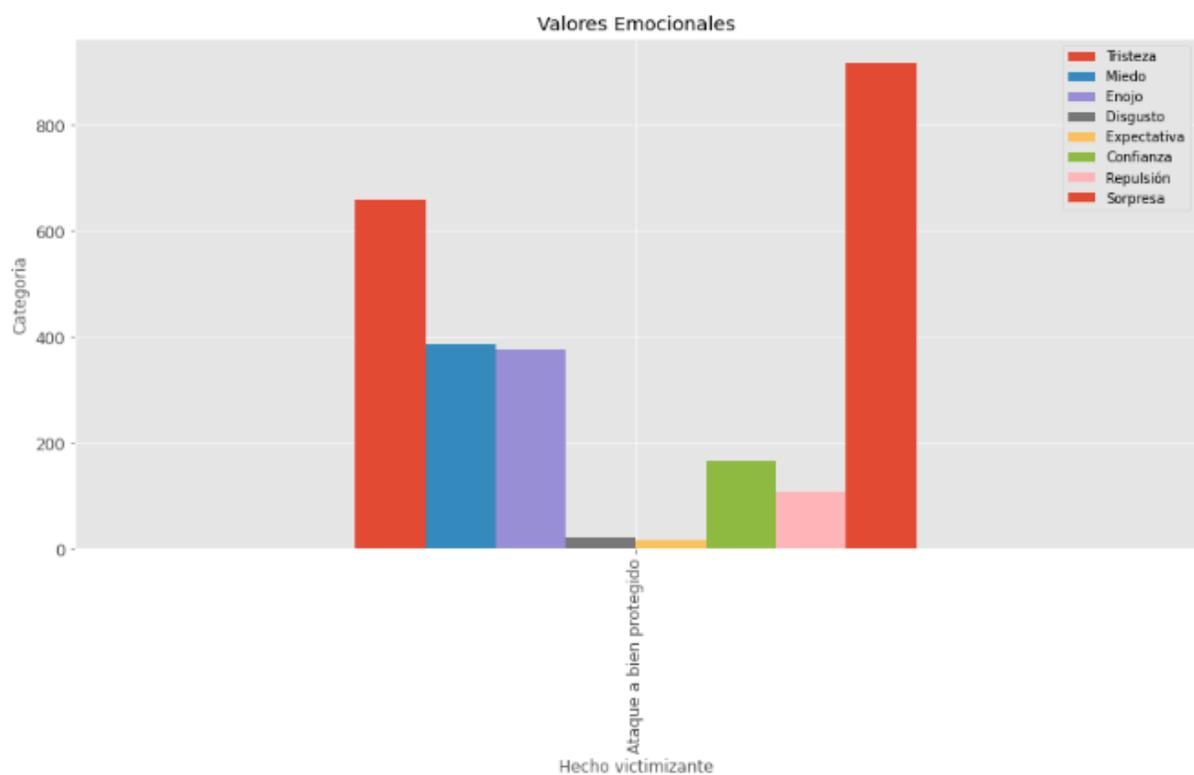


Figura 5-5: Categorías emocionales en “Ataque a bien protegido”

Tabla 5-5: Categorías emocionales en “Ataque a bien protegido”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Ataque a bien protegido	375	16	21	918	165	106	386	660

Para el último ejemplo se encuentran a disposición la imagen 5-6 y la tabla 5-6, las cuales arrojan niveles de reacción emocional muy altos, comparativamente, para desaparición encubridora (aproximadamente el 34% de las referencias afectivas para este macro hecho victimizante). La intensidad de la respuesta emocional, muy seguramente se debe a que el homicidio de la persona desaparecida es casi una certeza, pero no hay ni la confirmación, ni

la evidencia física para poder iniciar un duelo. Lo anterior en congruencia con la desaparición forzada: según al Artículo 165 del Código Penal, la desaparición forzada es cuando se “somete a otra persona a privación de su libertad cualquiera que sea la forma, seguida de su ocultamiento y de la negativa a reconocer dicha privación o de dar información sobre su paradero” [6].

Todos los hechos victimizantes agrupados en desaparición forzada muestran en común, los altos niveles de sorpresa, tristeza, miedo y enojo; mientras la confianza, reacción emocional que era de las menos comunes, figura esta vez con mayor frecuencia. Probablemente, esto se origine en que la desaparición forzada, en todos sus tipos, genera una situación de zozobra y preocupación por lo que en realidad haya ocurrido

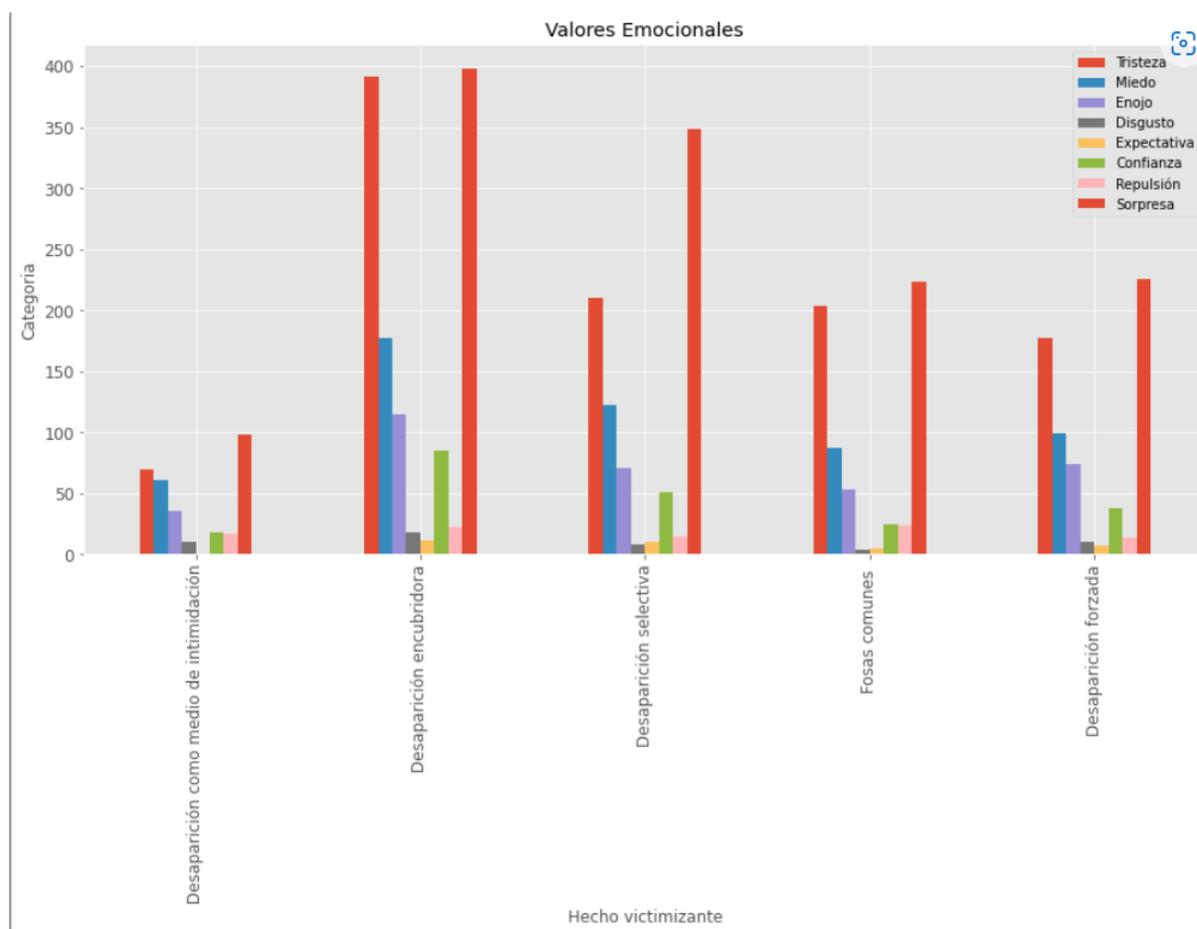
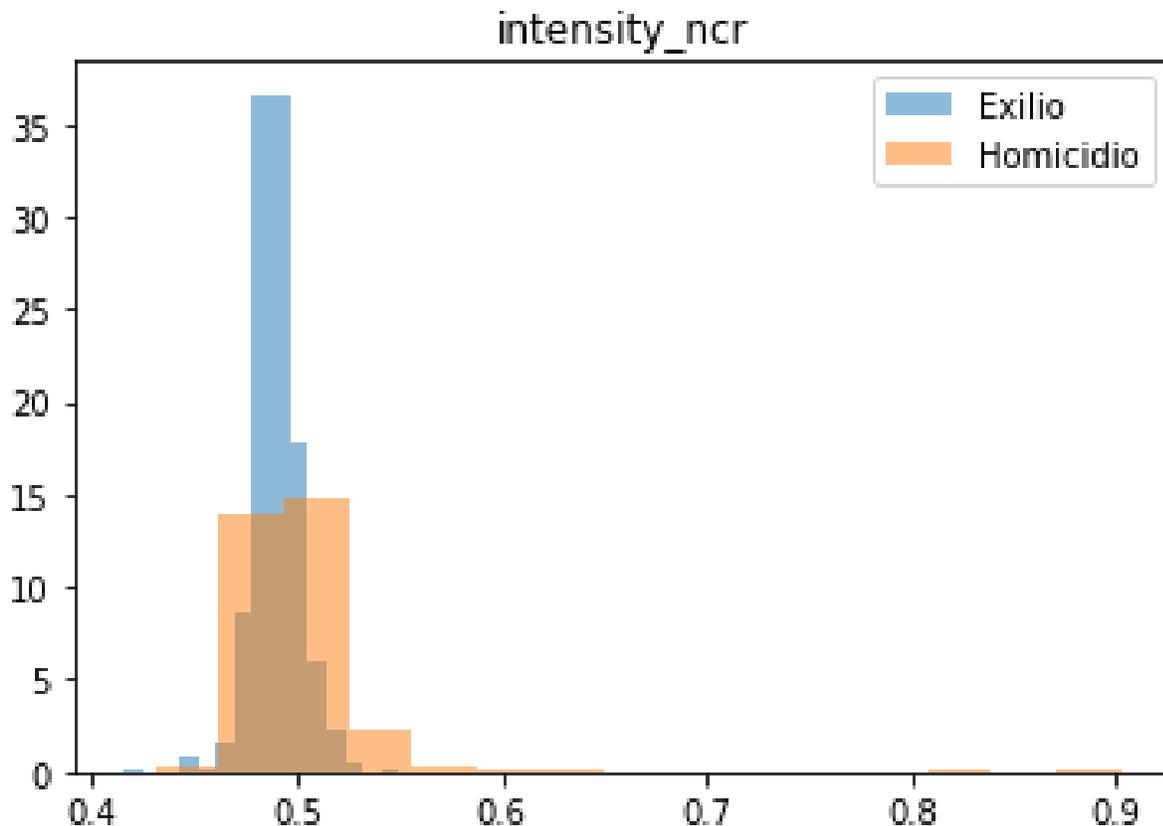


Figura 5-6: Categorías emocionales en “Desaparición forzada”

Tabla 5-6: Categorías emocionales en “Desaparición forzada”

	Enojo	Expectativa	Disgusto	Sorpresa	Confianza	Repulsión	Miedo	Tristeza
Desaparición como medio de intimidación	36	0	10	98	18	17	61	69
Desaparición encubridora	115	11	18	398	85	22	177	391
Desaparición selectiva	71	10	8	349	51	15	122	210
Fosas comunes	53	5	4	223	24	23	87	204
Desaparición forzada	74	7	10	226	38	13	99	177

Ahora bien, para un ejemplo de la comparación de intensidades entre dos categorías, se pueden apreciar las imágenes 5-7 y 5-8, las cuales analizan el par de categorías Exilio – Homicidio. Es evidente que, para el caso del primero, la intensidad se concentra en solamente un punto de la tabla (0,5; y 4-4.5 respectivamente). Esto habla de la especificidad de este hecho victimizante, y en comparación con otros hechos victimizantes analizados dentro de este informe, el promedio de su intensidad es más bajo. Por otro lado, se encuentra que el homicidio muestra intensidades más variables, pero que a la vez alcanza la mayor intensidad vista hasta el momento, con 0,9 de 1. Esto porque la pérdida de vida tiende a causar un impacto emocional alto, y más si se trata de una muerte violenta.

**Figura 5-7:** Intensity NCR en “Exilio-Homicidio”

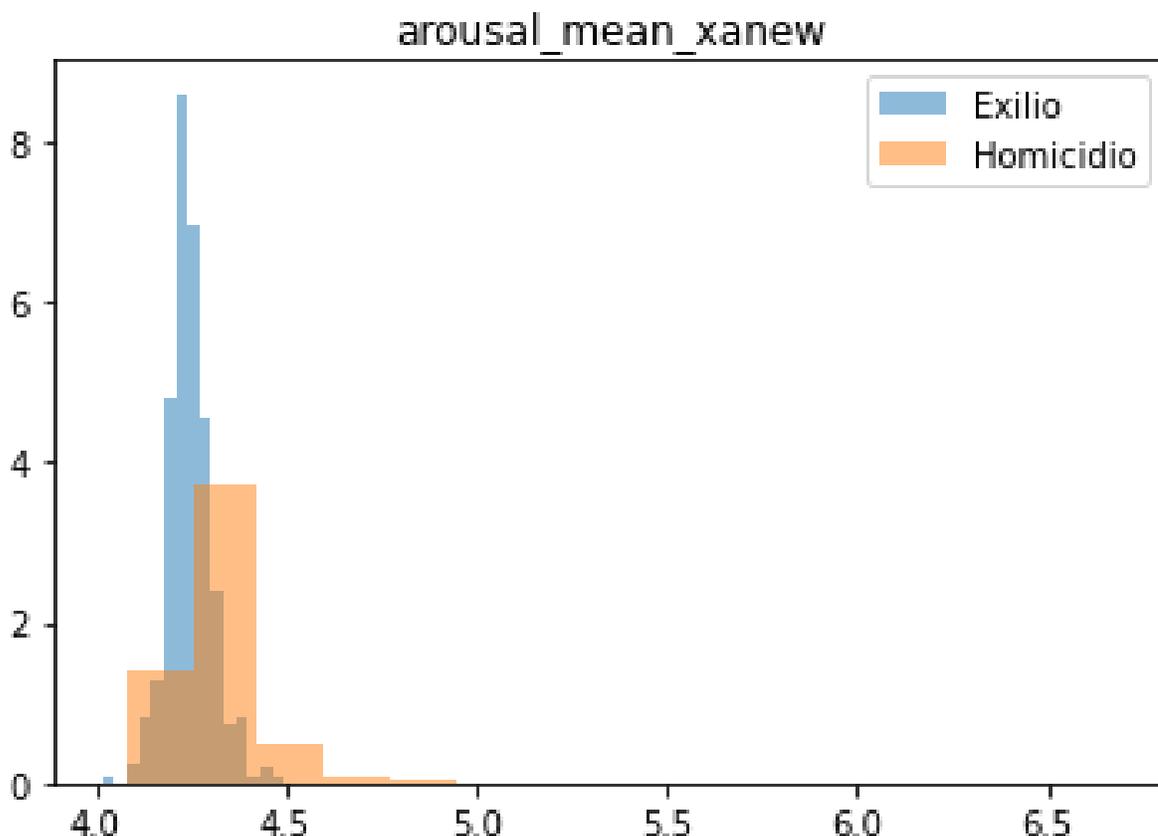


Figura 5-8: Arousal Anew en “Exilio-Homicidio”

Como conclusiones generales, es importante al hacer este tipo de análisis de reacciones emocionales, maximizar cuando la información fue obtenida a través de entrevistas, que no puede esperarse una traducción totalmente “objetiva” de respuestas que son eminentemente subjetivas. La percepción individual, la vivencia, y las palabras con que expresa un entrevistado su experiencia imprimen este componente subjetivo y por tanto hay que ser cuidadoso al momento de cuantificar y graficar este tipo de respuesta. A esta diferencia subjetiva entre víctimas, se puede sumar el preconceito subjetivo del observador externo, que en ocasiones puede pensar que hay hechos victimizantes más graves que otros, y si no se controla desde el comienzo el prejuicio, se corre el riesgo de matizar e interpretar con la distorsión del observador.

Esta misma subjetividad hace que en comparaciones como las numéricas, en cuanto a intensidades, sean similares, pues es una o varias personas quienes experimentan cierto evento, y aunque para otras personas pueda ser “menos grave”, para el que lo está viviendo es la totalidad de su experiencia. Lo que nuevamente hace problemático tratar de objetivar lo subjetivo.

Otra consideración, para valorar adecuadamente estos hallazgos, es que la información pro-

viene de un etiquetado con fragmentos de entrevistas variables en longitud y por tanto en número de palabras utilizadas, lo que podría variar artificialmente las frecuencias. Asimismo, como la cantidad de textos presentes dentro de cada etiqueta, pues, aunque se intentó equilibrar los corpus, en algunos casos específicos, no fue posible.

Es importante tener en cuenta, a modo de conclusión, que como se trata de textos narrativos, donde las personas cuentan sus experiencias personales, hay palabras propias del delito y su narración cronológica que pueden afectar los resultados, más allá que solo las emociones del entrevistado. Asimismo, como el uso del lenguaje coloquial, las palabras mal conjugadas según la lengua estándar o las variables fonéticas que son transcritas tal cual el entrevistado las pronuncia.

Un patrón general en todos los hechos victimizantes tratados en este texto, muestra una elevada frecuencia de la sorpresa y de la tristeza (cuando desde un posible prejuicio del observador, ciudadano y que no ha vivido el conflicto armado directamente podría pensarse que la mayor frecuencia sería la del miedo). Este predominio hace pensar que la inmediatez es una de las características más importante en los distintos *modus operandi* que se tienen cuando se ejecuta un hecho victimizante, lo que hace más fácil controlar a la víctima y generar incertidumbre; y la tristeza podría ser una consecuencia de ser víctima de eventos dolorosos y que cambian aspectos de la vida de quienes lo sufren.

Otra tendencia en común es la baja frecuencia del disgusto dentro de las entrevistas de las víctimas, mientras que el enojo figura en casi todos los hechos victimizantes (Así no sea la emoción más común). Este patrón podría deberse a que frente a hechos tan dolorosos para quien vivió esta experiencia, el disgusto es una emoción leve respecto a lo que las personas que vivieron directamente el conflicto armado experimentaron, y por eso se materializa más hacia el enojo.

Bibliografía

- [1] Código Penal Artículo 178, Ley 890 de 2004.
- [2] Código Penal Artículo 176, Ley 890 de 2004.
- [3] Jurisdicción Especial para la Paz, “La JEP hace pública la estrategia de priorización dentro del Caso 03, conocido como el de falsos positivos..”
- [4] Código Penal Artículo 180, Ley 890 de 2004.
- [5] I. Rodriguez, “Los bienes protegidos por el di: Bienes civiles, sanitarios, religiosos, culturales y obras e instalaciones que contengan fuerzas peligrosas,” s.f.
- [6] Código Penal Artículo 165, Ley 890 de 2004.

6 Frecuencias por temas y ministerios

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las entrevistas de los diferentes corpus utilizados dentro de los diversos informes que comprenden varios objetivos y preguntas de investigación.

En específico, la metodología de este capítulo se encamina bajo la directriz simple de tres pasos: 1. Obtención de un determinado corpus objetivo para su estudio, análisis y comprensión de resultados; 2. Procesamiento del corpus por medio de técnicas de trabajo de Procesamiento de Lenguaje Natural (PLN); 3. Análisis de resultados en función de los diferentes objetivos propuestos para cada informe, resaltando las directrices de categorías desarrolladas en el segundo paso.

Como aspecto metodológico es importante precisar el segundo paso, puesto que la gran problemática en el procesamiento de los datos se encuentra en este punto. Para ello, se trabaja con un código desarrollado en Google Colab donde en primera instancia se importan diferentes librerías como Spacy, Pandas, Os, Sys y Glob.

Spacy es una librería de software para procesamiento de lenguajes naturales desarrollado por Matt Honnibal y programado en lenguaje Python. Su utilidad en nuestra propuesta es la capacidad que otorga para tokenizar, es decir, dividir el texto en unidades lingüísticas individuales de manera automática [1]. De esta manera, Spacy nos permite reconocer la categoría gramatical de una palabra dada, además de filtrar su lema, lo que hace más fácil la identificación sin importar la flexión verbal en la que se encuentre.

Pandas es una librería para Python que se usa para el análisis de datos. Fue creada en el año 2008 por el programador Wes McKinney como una solución a la necesidad de tener una herramienta flexible y robusta para el análisis cuantitativo. Pandas está construido sobre 2 librerías clave de Python: matplotlib, que se usa para la visualización de datos, y NumPy, que ejecuta todas las operaciones matemáticas detrás.

Os es un módulo de la librería estándar de MicroPython que proporciona un conjunto de funciones para manipular los archivos y estructura de directorios, las funciones generales `uname` –da información del intérprete y/o microcontrolador– y `urandom` –genera bytes aleatorios–, funciones para el redireccionamiento y duplicación del terminal y funciones para la integra-

ción del sistema de archivos.

Sys es un módulo de la librería estándar de MicroPython que proporciona acceso a variables y funciones específicas del sistema. Son variables utilizadas o mantenidas por el intérprete y funciones que interactúan estrechamente con el intérprete.

Glob es una abreviatura de global command, una librería escrita en lenguaje C y muy utilizada en entornos Unix y GNU/Linux. Se diferencia fácilmente de las expresiones regulares por su sencillez y sus diferentes equivalencias en cuanto a los caracteres comodines utilizados y su extenso uso fuera de su entorno. Su código fuente es ampliamente publicado por pertenecer a Software libre y cualquier persona puede consultarlo. Si bien dicha función tiene estrictas reglas de uso en programación, su utilidad es fácilmente comprensible para la mayoría de los programadores al punto tal que forma parte intrínseca del shell de muchos sistemas operativos e incluso otros lenguajes como SQL en su Lenguaje de manipulación de datos (DML Data Manipulation Language), Pascal y Python.

La importación de estas entrevistas ayuda al procesamiento con el corpus, no obstante, no es el único componente por desarrollar, puesto que también se procesa un lexicón importado del interés acorde con la investigación en curso.

Por último, la presentación de los resultados se obtiene mediante el uso de tablas de frecuencias que denoten el uso de los diferentes términos de las diversas categorías en un periodo específico.

6.1. Análisis de los discursos presidenciales de posesión en Colombia según términos referentes a temáticas y ministerios

Este informe se culminó el 3 de febrero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. Este documento presenta el trabajo metodológico y analítico en torno al corpus de discursos presidenciales de posesión comprendidos en el período de mandato desde Gustavo Rojas Pinilla (1953-1957) hasta Iván Duque (2018-2022). La pregunta guía a abordar en este análisis viene a ser: ¿cómo los discursos presidenciales de posesión se pueden asociar con hechos y acciones puntuales en el mandato que cada presidente ha tenido? Así pues, luego de un trabajo de procesamiento del corpus por medio de un lexicón creado con términos y temas referentes a los 18 ministerios que existen hoy en día en Colombia, se procede a analizar cada discurso en función de los índices de frecuencia de términos asociados por temáticas y a su vez ministerios. El objetivo de este análisis es dar cuenta de tendencias y relaciones entre los discursos de posesión tanto con el trabajo realizado en el mandato de

cada presidente como con los hechos sociales más relevantes de cada período.

La obtención del corpus objetivo presentó ciertos obstáculos, puesto que en principio la idea era trabajar con los discursos presidenciales de la República de Colombia, lo que significaría empezar con el mandato del General José María Campo Serrano en 1886. No obstante, dicho trabajo resulta inasequible, ya que el registro digital de los discursos presidenciales de posesión que datan de ese año no existe. De tal modo, se inicia con el discurso del mandato de Gustavo Rojas Pinilla hasta el mandato de Iván Duque, así pues, los discursos presidenciales de posesión son obtenidos de repositorios institucionales, grabaciones de audio y documentos de prensa. Aclaración importante a esta etapa es que el discurso de Misael Pastrana Borrero no se logró encontrar, de tal modo se descarta su análisis.

En relación con el procesamiento del corpus, el lexicón realizado corresponde a una compilación de términos relacionados a los ministerios de Colombia, se hace revisión de descripciones, misión/visión, dependencias, nombres de ministros y glosarios. Toda la información es consultada en la página web oficial de cada ministerio. El lexicón está conformado por 2408 términos relacionados por temática (143 en total) y el nombre del ministerio al que pertenecen (18 ministerios).

Con los resultados obtenidos del procesamiento en donde se tiene por cada discurso una gráfica que da cuenta de los índices de frecuencia de palabras asociadas a ministerios y a su vez otra gráfica con el índice de términos asociados por temáticas se trabaja entonces en la redacción de información donde se señalan los índices con mayor y menor frecuencia, a su vez como una propuesta de interpretación e hipótesis se da un contexto global de cómo cada mandatario asume el cargo de presidencia y también un relacionamiento con los hechos y acciones principales llevadas a cabo durante cada mandato.

A continuación, se presentan dos ejemplos de los discursos presidenciales de dos presidentes en específico. El primero es Álvaro Uribe Vélez en su primer periodo presidencial y el segundo es Juan Manuel Santos en su primer periodo presidencial. Para el primer mandato de Uribe se obtienen las imágenes **6-1** y **6-2**.

Álvaro Uribe Vélez es un abogado especializado en administración y gerencia, que llegó a la presidencia, en su primer periodo, entre 2002 y 2006 avalado por el movimiento Primero Colombia bajo el discurso de la seguridad democrática y la lucha contra el narcotráfico y terrorismo. Antes de su primer periodo presidencial, desempeñó varios cargos como secretario general del Ministerio del Trabajo en 1977, director de la Aeronáutica Civil en 1980, alcalde de Medellín en 1982, concejal y senador entre 1984 y 1986 y posteriormente gobernador de Antioquia en 1995 [2].

Ahora bien, en relación con la información de las imágenes, la primera de ellas muestra el predominio de términos relacionados con el Ministerio de Defensa como libertad, lucha,

autoridad, seguridad, entre otros. Posteriormente, son recurrentes el Ministerio de interior y el de Justicia del Derecho. La segunda imagen muestra las temáticas del discurso en cuestión. Aquí se encuentra la tendencia principal hacia las normas de convivencia, y seguidamente de problemáticas de defensa, gobierno y sociedad.

A partir de este análisis del discurso, es evidente, en primera instancia, que sus palabras concuerdan con su discurso como candidato y su recorrido político previo, como la promoción de estrategias de seguridad privada tales como Las Convivir en su labor como Gobernador de Antioquia. Esto está relacionado directamente con el sonoro fracaso del proceso de paz intentado durante el gobierno de Andrés Pastrana entre 1998 y 2002, en el cual la sensación de impunidad generalizada y del surgimiento de un Estado paralelo dentro de las fronteras del país, generó un ambiente nacional propicio a las políticas de mayor rigor, el sometimiento a la autoridad presidencial y restablecimiento del orden público por medio de las fuerzas armadas [3].

Dentro de este contexto, es lógico que el predominio del contenido discursivo se enfoque en labores propias del Ministerio de Defensa. También resulta coherente que las temáticas de normas de convivencia sean abordadas, pues en un país que se percibía como completamente anómico, era compatible con el deseo popular de la recuperación del acatamiento a la autoridad y las ansias de seguridad jurídica. Por eso mismo la presencia del Ministerio de Interior, pues toma relevancia porque en el imaginario colectivo se había perdido la fuerza de la figura presidencial, característica de la historia constitucional colombiana (análisis que se enlazan con los resultados de la gráfica de temáticas). Por lo que fortalecer este Ministerio permitiría afianzar la autoridad presidencial a través de políticas más autoritarias y enfocadas más en el orden público.

Del mismo modo, Álvaro Uribe logró la reelección tras la aprobación del acto legislativo 02 de 2004, que permitió la reelección inmediata del presidente. Después de ser electo en su primer periodo con mayoría absoluta y sin necesidad de segunda vuelta, con un 53,05 %, [4]; y tras haber alcanzado una aprobación popular del 80 % [5], vuelve a ganar con mayoría absoluta con 61,23 % de los votos [6].

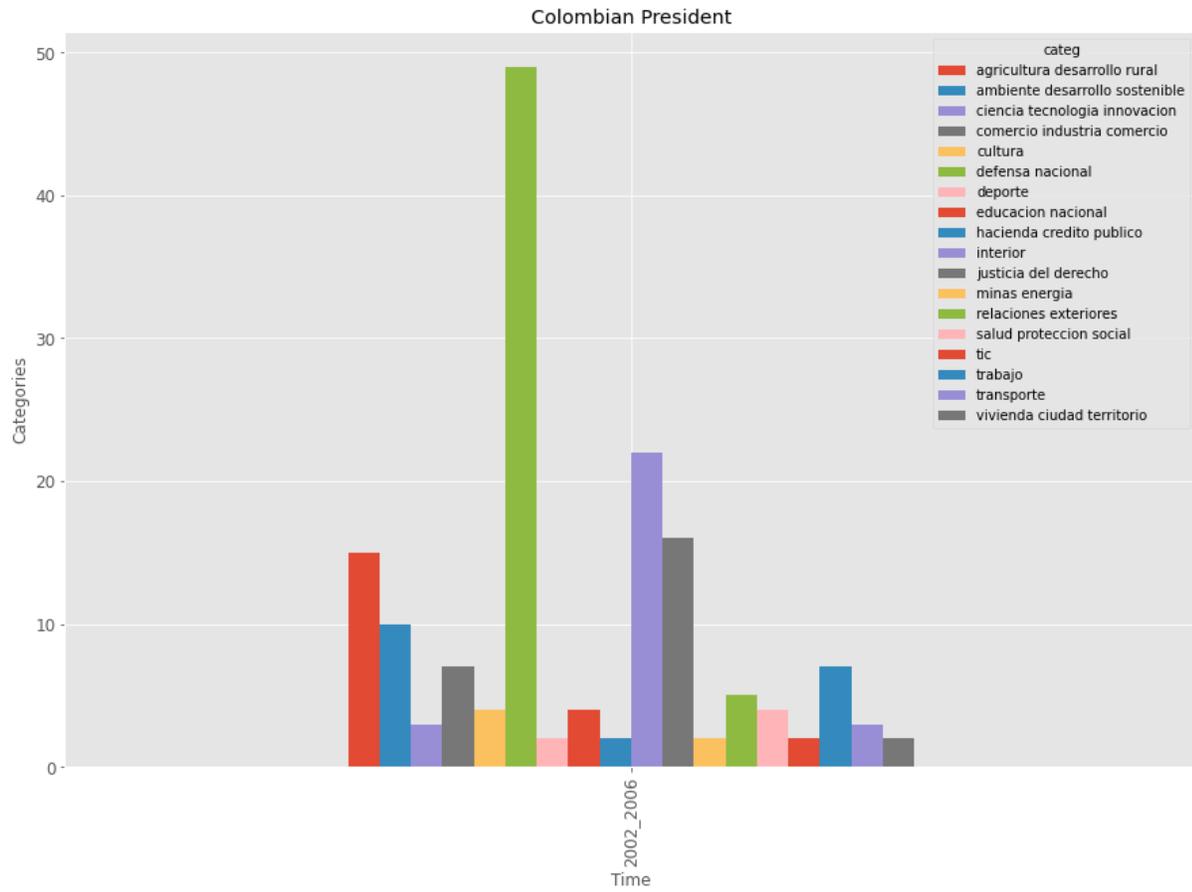


Figura 6-1: Álvaro Uribe (2002-2006) - Frecuencia de términos por ministerio en el discurso

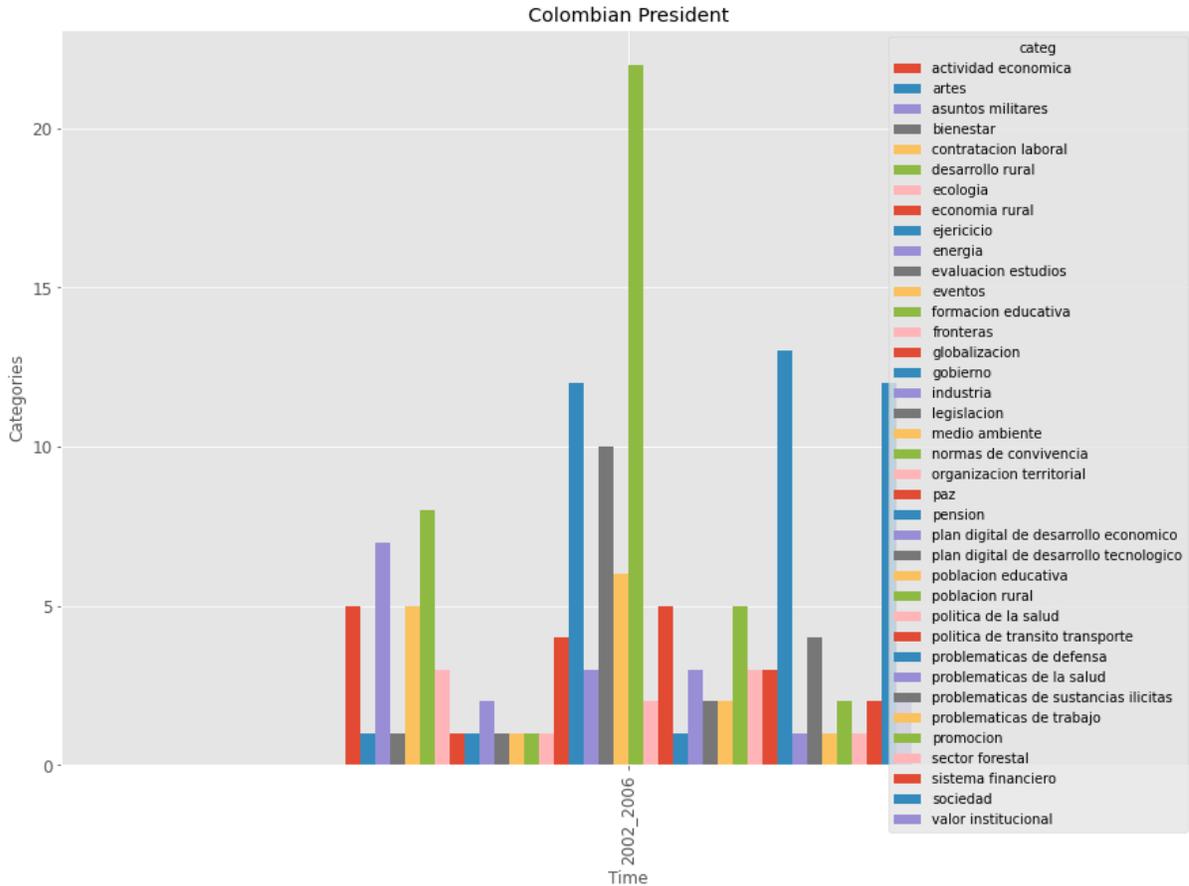


Figura 6-2: Álvaro Uribe (2002-2006) - Frecuencia de términos por tema en el discurso

Ahora bien, para el caso del primer periodo presidencial de Juan Manuel Santos, se obtienen las imágenes **6-3** y **6-4**.

Juan Manuel Santos es un político y economista que alcanzó la presidencia por el Partido de la U, después de que la Corte Constitucional declarara inexecutable el referendo que buscaba la segunda reelección de Uribe. Santos gozaba de la imagen de continuidad respecto del mandato de Álvaro Uribe Vélez [7]. Durante el periodo presidencial anterior ejerció como ministro de Defensa. [8].

Ahora bien, en relación con la información de las imágenes, la primera de ellas arroja la relevancia que tiene el Ministerio de Defensa Nacional, con términos como nación, gobierno y seguridad en su práctica discursiva. Adicionalmente hay una tendencia al uso de palabras como diversidad, tierra y campesino, propias del Ministerio de Agricultura y Desarrollo Rural. En consecuencia, la segunda imagen gira en torno, principalmente, a temas como gobierno, normas de convivencia, desarrollo y población rurales.

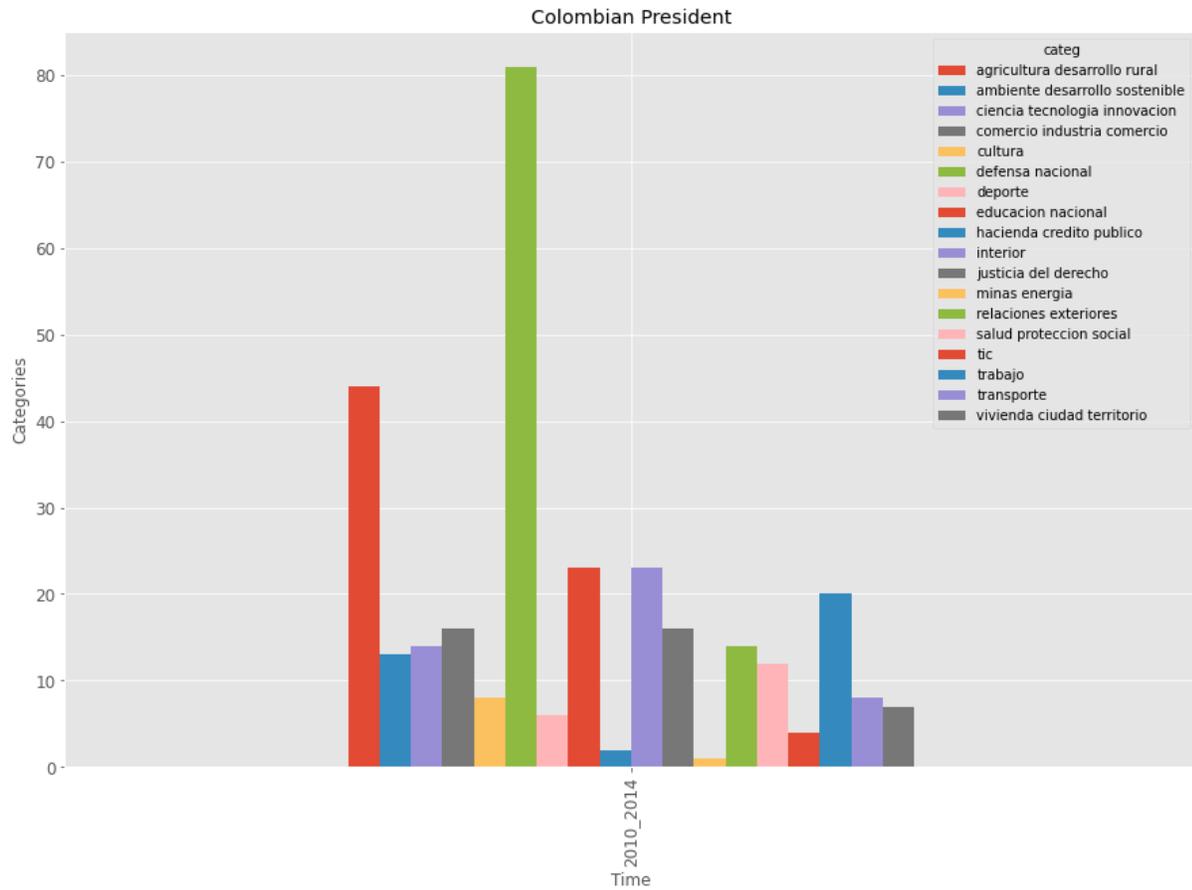


Figura 6-3: Juan Manuel Santos (2010-2014) - Frecuencia de términos por ministerio en el discurso

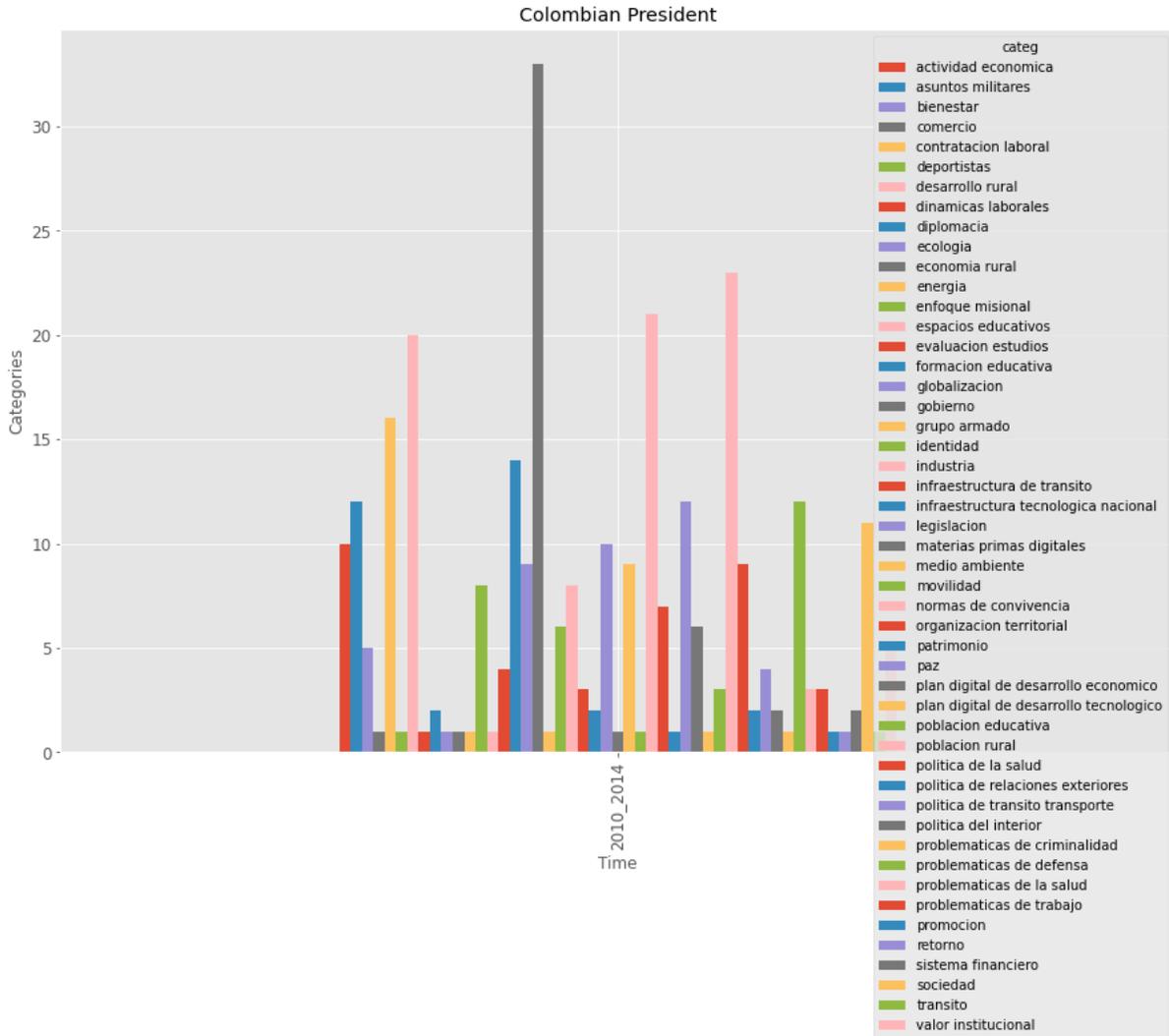


Figura 6-4: Juan Manuel Santos (2010-2014) - Frecuencia de términos por tema en el discurso

Dado que la candidatura de Juan Manuel Santos provino de que se le percibía como el continuador de la labor de Uribe Vélez y de su éxito como ministro de Defensa (bajo su ministerio se realizaron exitosas operaciones militares, entre ellas la memorada Operación Jaque y las bajas de líderes guerrilleros) se presumen la necesidad de mantener la consistencia discursiva, lo que se podría evidenciar en las temáticas de gobierno y normas de convivencia, y en la alta presencia del Ministerio de Defensa.

Por otra parte, puede notarse en las gráficas como tópicos relevantes tanto el Ministerio de Agricultura como las temáticas de desarrollo y población rural. Esto se comprende cuando se tiene en cuenta que, para el segundo periodo de Álvaro Uribe, las cifras de desplazamiento de población rural se habían incrementado a niveles previamente desconocidos: 2.4 millones de desplazados de los 4.9 millones entre 1985 y 2009 [9]. Por lo que era necesario buscar una

solución a esta problemática, apostándole de forma paralela al apoyo a la inversión extranjera y la tecnificación del campo, concordante con la recuperación de territorios previamente ocupados por las guerrillas y con la continuidad del TLC entre Colombia y Estados Unidos, que se venía gestando desde el gobierno anterior.

A partir del trabajo de procesamiento realizado al corpus de discursos presidenciales de posesión en un periodo aproximado de 70 años, se hace posible un análisis descriptivo e histórico al relacionar los datos de frecuencia de términos referentes a ministerios y sus temáticas con hechos ocurridos en cada mandato, de modo que se llega a las siguientes conclusiones:

El enfoque discursivo hegemónico que ha preponderado en la política colombiana, por ya casi un siglo, gira entorno a términos concretos como violencia, seguridad social y defensa.

El Ministerio de Defensa es el ministerio preponderante en los discursos presidenciales colombianos hasta la fecha.

En todos los discursos, el Ministerio de Defensa es el que presenta mayor número de apariciones, preponderando mayormente en el discurso de Alberto Lleras y disminuyendo hasta el discurso de Andrés Pastrana, donde la frecuencia nuevamente empieza a aumentar.

En todos los discursos, los términos asociados a Ministerios de Salud y de Educación presentan un bajo índice de frecuencia.

Como nota aclaratoria y como se señala en el apartado de metodología, el discurso de posesión de Misael Pastrana no fue posible encontrarlo. Se informa que su búsqueda se hizo en páginas institucionales, repositorios públicos y por último en la biblioteca de Señal Memoria de RTVC, donde se encontró el material audiovisual del discurso, pero al hacer su revisión nos encontramos con una grabación de una hora de duración en la cual se presentan las palabras preliminares al discurso y posteriormente un archivo de video con solamente un sonido de estática.

A través del análisis terminológico, se posibilita a investigadores de distintas dependencias aproximarse a un trabajo de análisis descriptivo con un soporte más concreto de datos, de tal manera que se invita a profundizar en temas clave como viene a ser el relacionamiento de hechos históricos con el término y su frecuencia en los discursos de posesión.

Bibliografía

- [1] Oxford, "Oxford languages," 2021.
- [2] Álvaro Uribe Vélez, "Carrera política," 2019.
- [3] E. Paulo, "La incidencia del fracaso del proceso de paz entre el gobierno nacional y las fuerzas armadas revolucionarias de colombia (farc-ep) en la implementación de la política pública de seguridad democrática en colombia, durante el período comprendido entre 1999 y 2006," monography, Universidad Colegio Mayor de Nuestra Señora del Rosario, Bogota D.C., Colombia, 2011.
- [4] M. C. Osorio, "Primer mandato presidencial de uribe vélez," Master's thesis, Universidad Andina Simón Bolívar, 2012.
- [5] R. E. País, "Presidente Álvaro uribe termina su gestión con 80 % de aprobación," Jul. 30 2010.
- [6] Georgetown University, "República de colombia resultados electorales 2006," Mar. 12 2006.
- [7] P. Lozano, "El constitucional colombiano cierra la vía a un tercer mandato de uribe," Feb. 26, 2010.
- [8] Presidencia de la República de Colombia, "Palabras del presidente juan manuel santos en su posesión para el período presidencial 2014-2018," 2014.
- [9] I. Coello, "El gobierno uribe es el que más personas ha desplazado," Ago. 8 2010.

7 Sistema de recomendación

La metodología desarrollada en este capítulo comprende el análisis del procesamiento de las entrevistas de los diferentes corpus utilizados dentro de los diversos informes que comprenden múltiples objetivos y preguntas de investigación.

El trabajo metodológico se fundamenta en el Procesamiento de lenguaje Natural de las entrevistas objetivo dentro de un sistema de recomendación. Este sistema busca presentar recomendaciones para que la persona que este leyendo la entrevista pueda tener un contexto más amplio de la situación mediante la sugerencia de otras entrevistas. Cada entrevista tiene implementado un tesaurus, el cual contiene un árbol de etiquetas para etiquetar la entrevista bajo los fenómenos sociales que la Comisión de la Verdad desea analizar y otorgar una desagregación de los fenómenos sociales a partir de sucesos específicos que están enmarcados en los fenómenos sociales por analizar; por ello, la sugerencia de otras entrevistas se basa en encontrar una relación alta entre las etiquetas presentes en la entrevista que el lector lee y las etiquetas que están presentes en las restantes entrevistas que el corpus de la Comisión ha recopilado. Tales recomendaciones presentan las relaciones significativas de las entrevistas demostrando de forma más clara los distintos patrones y tendencias que se establecen dentro y entre las etiquetas.

Para el desarrollo de la metodología se deben seguir unos pasos preestablecidos:

1. **Recolección de información:** La plataforma sim3 de la comisión de la verdad tiene un componente propio donde se puede descargar la base de datos completa, por lo cual, el archivo descargado en un formato separado por comas (.csv) puede ser explorado en distintas herramientas como Excel u otros lenguajes de programación como Python. Su primer campo (columna) es un identificador de la entrevista y los restantes campos son todas las etiquetas diseñadas en el árbol de etiquetas. Con esto esto en mente, cada registro de la base de datos contiene la frecuencia de la etiqueta en la entrevista de modo que una entrevista puede tener múltiples menciones de la misma etiqueta.
2. **Filtración de la información:** El sistema establece los fenómenos sociales o suceso específico a estudiar de modo que las etiquetas específicas correspondientes a dichos fenómenos y/o sucesos para que sean las únicas presentes en esta nueva base de datos. De lo anterior, hay que resaltar que un fenómeno social puede tener muchas etiquetas

relacionadas causando que el sistema explore el árbol de etiquetas que se encuentre en la plataforma ya mencionada para tener los sucesos específicos a investigar.

3. Generación de frecuencias: Una vez obtenida la nueva base de datos del componente anterior, se dispone a transformar cada registro de la base de datos. La base contiene las frecuencias de las etiquetas en cada entrevista; por lo tanto, las frecuencias de las etiquetas serán modificadas para obtener una variable dicotómica que identifique la existencia de dicha etiqueta en la entrevista.
4. Creación de reglas de asociación: La creación de las reglas de asociación inicia implementando un algoritmo apriori a la base de datos descrita en el componente anterior. El algoritmo apriori es una técnica de minería de datos para identificar las asociaciones relevantes bajo tres métricas de evaluación esenciales: lift, confidence y support. Para la implementación de este algoritmo, se establece un valor mínimo en la métrica “support” donde el criterio del investigador es el principal responsable por colocar este valor. Posteriormente, se crea las reglas de asociación colocando un mínimo valor de la métrica de lift donde cada recomendación muestra sus valores en las tres métricas de evaluación ya mencionadas. A diferencia del anterior, este valor se busca que sea un valor cercano a 1.
5. Filtración de reglas basadas en condiciones: Las reglas basadas en condiciones son normas de adición donde el sistema muestra aquellas recomendaciones que cumplan ambas métricas de “lift” y “confidence”. Por lo tanto, el sistema busca ajustar los valores de ambas métricas con el objetivo de obtener unas recomendaciones significativas; es decir, unos valores en las métricas cercanas a 1. Ahora, la cantidad de los resultados en este componente pueden variar llevando al sistema a suavizar los valores de las métricas en aras de tener suficientes resultados para el sistema.
6. Recomendaciones válidas e inválidas: Dado la integración de diferentes dominios, este componente valida las recomendaciones encontrando etiquetas pertenecientes entre distintos dominios. Es decir, cuando el sistema es alimentado con un solo dominio, se entiende que todas las recomendaciones son válidas; por otro lado, cuando el sistema es alimentado con una base de datos con etiquetas pertenecientes a diferentes dominios, el lector asocia distintos sucesos que pertenecen a un fenómeno social distinto llegando al objetivo de otorgar un contexto más amplio.
7. Grafo de las recomendaciones: Este componente realiza los grafos sobre las recomendaciones que fueron validadas en el componente anterior. Los grafos están diseñados para tener como nodos (círculos) las etiquetas y las recomendaciones con el objetivo de enlazar las etiquetas que están relacionadas. Dichas relaciones son clasificadas entre dos tipos, enmarcando una relación unidireccional o bidireccional. Una relación unidireccional implica que la búsqueda de una entrevista con una etiqueta particular y

el sistema sugiere otra etiqueta presente en otra entrevista. En cambio, una relación bidireccional implica que el sistema de recomendación sugiere al lector buscar uno de los nodos de la relación bidireccional cuando el otro nodo fue buscado; al igual, la sugerencia sería similar si los roles de búsqueda en los nodos son intercambiados. Los grafos presentan una relación bidireccional como un circuito cerrado entre dos nodos representantes de las etiquetas relacionadas y dos nodos representantes a una recomendación distinta. En contraste, una relación unidireccional tiene solo un sentido entre los nodos representantes de las etiquetas; es decir, hay un nodo que conecta a una recomendación y esta última a la etiqueta relacionada.

7.1. Análisis de sistema de recomendación de entrevistas etiquetadas por el equipo de resistencia

Este informe fue culminado el 4 de febrero de 2022 por parte del Grupo de Procesamiento de Lenguaje Natural. El informe tiene por finalidad exponer el estudio generado en base al corpus del texto destinado para el análisis del Sistema de recomendaciones presentes en entrevistas, el cual se encuentra compuesto por las entrevistas etiquetadas por el Equipo de Resistencia y por las entrevistas catalogadas bajo otras categorías de etiquetas tales como las de Hechos victimizantes, los afrontamientos y resistencias no violentas o cotidianas y la categoría de Resistencias organizadas. Esto con el objetivo de encontrar patrones generales en formas de reglas que posibiliten el desarrollo de la hipótesis en relación con estos temas para que sus resultados sirvan como aporte y base para futuros procesos de investigación que ayuden al esclarecimiento y entendimiento de los eventos sucedidos en el marco del conflicto armado en Colombia.

A continuación, se presentarán los resultados de la metodología aplicada. Cada etiqueta de interés tiene una subsección individual describiendo los aspectos sociales relevantes con el fin de relacionar los resultados sugeridos del sistema de recomendación. Por lo tanto, los resultados contienen 4 apartados asociando el estudio de las etiquetas de interés a analizar:

- Resistencias organizadas
- Resistencias organizadas vs Afrontamientos y resistencias no violentas – cotidianas
- Resistencias organizadas vs Hechos victimizantes
- Resistencias organizadas, Afrontamientos y resistencias no violentas – cotidianas y Hechos victimizantes

A continuación, se presentarán los resultados de resistencias organizadas. Se busca demostrar las relaciones entre las etiquetas de menor nivel que conforma la etiqueta mayor de resistencias organizadas. La tabla **7-1** muestra las seis recomendaciones producto del sistema de

recomendación, donde el lift es la métrica de evaluación del sistema a variar para identificar sus cambios. En consecuencia, en esta ocasión se establece un lift de 1.

Tabla 7-1: Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=1

Antecedentes	Consecuencias	Support antecedentes	Support consecuencias	Support	Confidence	Lift	Leverage	Conviction
Acciones formativas	Campesinado	0.094945	0.172010	0.017263	0.181818	1.057022	0.000931	1.011988
Campesinado	Acciones formativas	0.172010	0.094945	0.017263	0.100358	1.057022	0.000931	1.006018
Apoyo legal	Apoyo psicosocial	0.086930	0.147349	0.016646	0.191489	1.299564	0.003837	1.054595
Apoyo psicosocial	Apoyo legal	0.147349	0.086930	0.016646	0.112971	1.299564	0.003837	1.029358
Acciones formativas	Resistencia	0.094945	0.173859	0.019729	0.207792	1.195174	0.003222	1.042833
Resistencia	Acciones formativas	0.173859	0.094945	0.019729	0.113475	1.195174	0.003222	1.020903

A su vez, la imagen 7-1 establece las relaciones existentes entre las entrevistas de resistencia. Como se puede evidenciar, en la imagen se establecieron 6 reglas en donde se evidencian tres relaciones bidireccionales, entre estas se encuentran: las acciones formativas y la resistencia, las acciones formativas y el campesinado, y una última relación entre el apoyo psico-social y el apoyo legal.

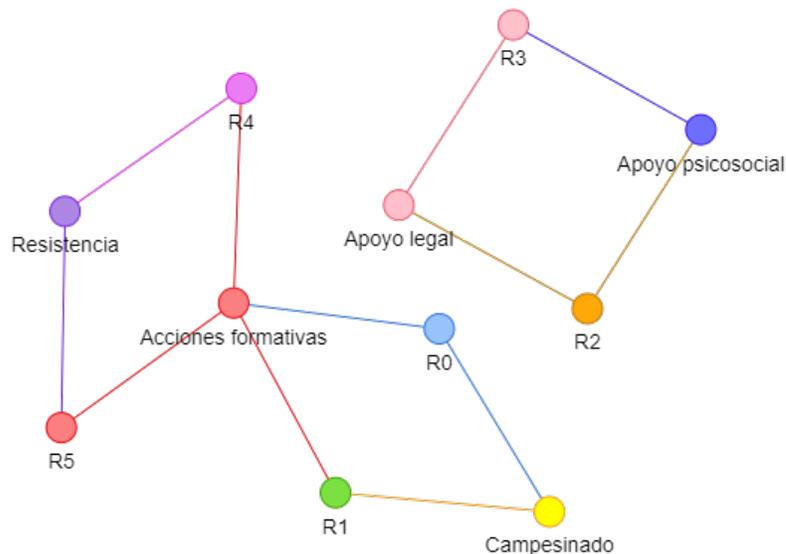


Figura 7-1: Grafo de relaciones para Resistencia n=6, lift=1, confidence=0.1

En relación a la primera regla mencionada (R4 y R5), se puede inferir que los actos de resistencia requieren de acciones formativas específicas, estas pueden ser formas de expresión

cultural, artística y política que inspiran y le dan forma al tipo de resistencia que se quiere llevar a cabo. Al respecto, también se pueden encontrar relaciones conmensurables entre las distintas formas de expresión, esto quiere decir, que las expresiones pueden ser compartidas y sus acciones formativas en común son las que le dan la forma a los actos de resistencia. Un ejemplo de esto ocurre en los actos de resistencia durante la dictadura de Brasil iniciada en 1964. Esto se debe a que estos actos fueron producto de acciones formativas en los ámbitos culturales y políticos. “Político, porque ayudó en el proceso de reorganización de los partidos de izquierda y en la revisión de los postulados ideológicos del Partido Comunista Brasileño. Y cultural, porque esa reorganización se dio muchas veces en el ámbito de las producciones culturales, en el cual las izquierdas constituyeron un espacio de cuestionamiento y de compromiso a través de las artes y las actividades intelectuales” [1].

En la siguiente relación (R0 Y R1) se encuentra a un actor específico, que es el campesinado cuyo rol dentro del sistema es llevar a cabo las acciones formativas. La explicación de esta relación dentro de la recomendación con el lift proporcionado surge con el reconocimiento del sector campesino como uno de los más afectados por el conflicto armado de acuerdo con la Comisión de la Verdad. No es únicamente un sector de víctimas, sino que también posee agencia dentro de la resistencia al conflicto armado al interior del país: “El campesinado ha sido, realmente, parte central de la columna vertebral de las resistencias a la guerra y de la construcción de la paz” [2].

Finalmente, en la última recomendación (R2 y R3), se evidencia que se puede encontrar información relevante cuando se relacionan las entrevistas sobre el apoyo legal y el apoyo psicosocial, esto debido a que un elemento relevante dentro de la construcción de resistencias se basa en el conocimiento y aplicación de saberes en elementos tanto legales como psicosociales para hacer que tanto la comunicación como el bienestar de los grupos, personas y organizaciones que resisten tengan la estabilidad necesaria.

Por otro lado, la tabla 7-2 presenta los resultados del sistema de recomendación cuando la métrica ya mencionada se cambia por un valor inferior de 0.6. Es así que, para denotar un cambio sustancial sobre las recomendaciones, el lift fue cambiado por 0.4 unidades para describir las relaciones adicionales.

Tabla 7-2: Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=0.6

Antecedentes	Consecuencias	Support antecedentes	Support consecuencias	Support	Confidence	Lift	Leverage	Conviction
Construcción paz	Estrategias resistencia	0.080764	0.139334	0.010481	0.129771	0.931365	-0.000772	0.989011
Construcción paz	Resistencia	0.080764	0.173859	0.011714	0.145038	0.834227	-0.002328	0.966289
Acciones formativas	Campesinado	0.094945	0.172010	0.017263	0.181818	1.057022	0.000931	1.011988

Tabla 7-2: Resultados del sistema de recomendación para las entrevistas resistencias con la etiqueta de Resistencia para un lift=0.6

Antecedentes	Consecuencias	Support antecedentes	Support consecuencias	Support	Confidence	Lift	Leverage	Conviction
Campesinado	Acciones formativas	0.172010	0.094945	0.017263	0.100358	1.057022	0.000931	1.006018
Campesinado	Resistencia	0.172010	0.173859	0.025894	0.150538	0.865858	-0.004012	0.972545
Resistencia	Campesinado	0.173859	0.172010	0.025894	0.148936	0.865858	-0.004012	0.972888
Organizaciones víctimas	Apoyo psicosocial	0.356967	0.147349	0.046856	0.131261	0.890816	-0.005743	0.981481
Apoyo psicosocial	Organizaciones víctimas	0.147349	0.356967	0.046856	0.317992	0.890816	-0.005743	0.942852
Apoyo legal	Organizaciones víctimas	0.086930	0.356967	0.022195	0.255319	0.715246	-0.008836	0.863502
Artísticos deportivos	Organizaciones víctimas	0.043773	0.356967	0.013564	0.309859	0.868034	-0.002062	0.931742
Acciones formativas	Organizaciones víctimas	0.094945	0.356967	0.024044	0.253247	0.709441	-0.009848	0.861105
Apoyo legal	Apoyo psicosocial	0.086930	0.147349	0.016646	0.191489	1.299564	0.003837	1.054595
Apoyo psicosocial	Apoyo legal	0.147349	0.086930	0.016646	0.112971	1.299564	0.003837	1.029358
Apoyo legal	Resistencia	0.086930	0.173859	0.013564	0.156028	0.897440	-0.001550	0.978872
Acciones formativas	Resistencia	0.094945	0.173859	0.019729	0.207792	1.195174	0.003222	1.042833
Resistencia	Acciones formativas	0.173859	0.094945	0.019729	0.113475	1.195174	0.003222	1.020903

Por último, en la imagen 7-2 se establecen las relaciones existentes entre las entrevistas de resistencia, en la cual se pueden evidenciar 16 reglas pertenecientes a las recomendaciones. En cambio, en este grafo se integra un nodo que establece nuevas reglas de recomendación, este corresponde al nodo de Organizaciones de víctimas estableciendo 4 reglas dentro del sistema, una de carácter bidireccional y las otras unidireccionales. En la relación de recomendación bidireccional, se establece entre las organizaciones de víctimas y el apoyo psicosocial, si se revisan entrevistas en las que se encuentre información sobre esto se puede inferir que se relata sobre cómo las organizaciones de víctimas han establecido herramientas de apoyo psicosocial para su funcionamiento en objetivos específicos, elemento que también podría ser evidenciado si se analiza la regla 10 que une las organizaciones de víctimas con las acciones formativas. Retomando el aspecto relacional entre las organizaciones y víctimas y el apoyo psicosocial, es importante mencionar proyectos como el desarrollado por el ministerio de salud en relación con los apoyos en el marco del conflicto armado: “Procesos articulados de servicios que buscan mitigar, superar y prevenir los daños e impactos a la integridad psicológica y moral, al proyecto de vida y la vida en relación, generados a las víctimas, sus familias y comunidades por las graves violaciones de Derechos Humanos e infracciones al Derecho Internacional Humanitario. La atención psicosocial de las víctimas en el marco de la reparación tiene tres modalidades: Individual, familiar y comunitaria; y se desarrollará en cuatro fases: focalización y contacto, caracterización, atención y cierre” [3].

Las otras dos relaciones establecidas en relación con este nodo son la relación 9 que une el nodo de Organizaciones de víctimas con el de artísticos deportivos, una relación de recomendación interesante para el análisis por ver si este tipo de organizaciones tienen este carácter. La última relación establecida es entre este nodo y apoyo legal. Esta relación contiene varias vertientes, puesto que el ámbito legal puede ser distribuido a varias causales. Se pueden interponer dos ejemplos claros. El primero de ellos es que “La Unidad de Víctimas

hace acompañamiento técnico a las Mesas de Participación Efectiva de Víctimas que operan a nivel municipal, departamental y nacional, para que estos escenarios operen de acuerdo a lo dispuesto en el Protocolo de Participación Efectiva de las Víctimas expedido por medio de la Resolución 388 de 2013” [4]. El segundo de ellos es la ley 1448 de 2011 o ley de víctimas y restitución de tierras: “...fue creada con el fin de garantizar el derecho de las víctimas. De este modo, esta política pública se enmarca en los derechos de las víctimas, en reconocer la existencia de un conflicto armado en el país y en reconocer a las víctimas sin importar su victimario. Bajo este escenario se direccionan una serie de medidas individuales y colectivas, que apuntan a la búsqueda de la verdad, justicia y reparación” [5].

En lo que concierne a la relación 0, se establece que si el investigador revisa las entrevistas relacionadas con las estrategias de resistencia, debería revisar las entrevistas en las que se habla sobre la construcción de paz. Esto puede ser debido a que uno de los objetivos principales de las estrategias de resistencia es asegurar la construcción de la paz. Por otro lado, en esta imagen se establece que hay una relación cercana entre resistencia y campesinado, cosa que no se establecía de forma directa en la gráfica anterior, al igual que la recomendación de la construcción de paz con la resistencia.

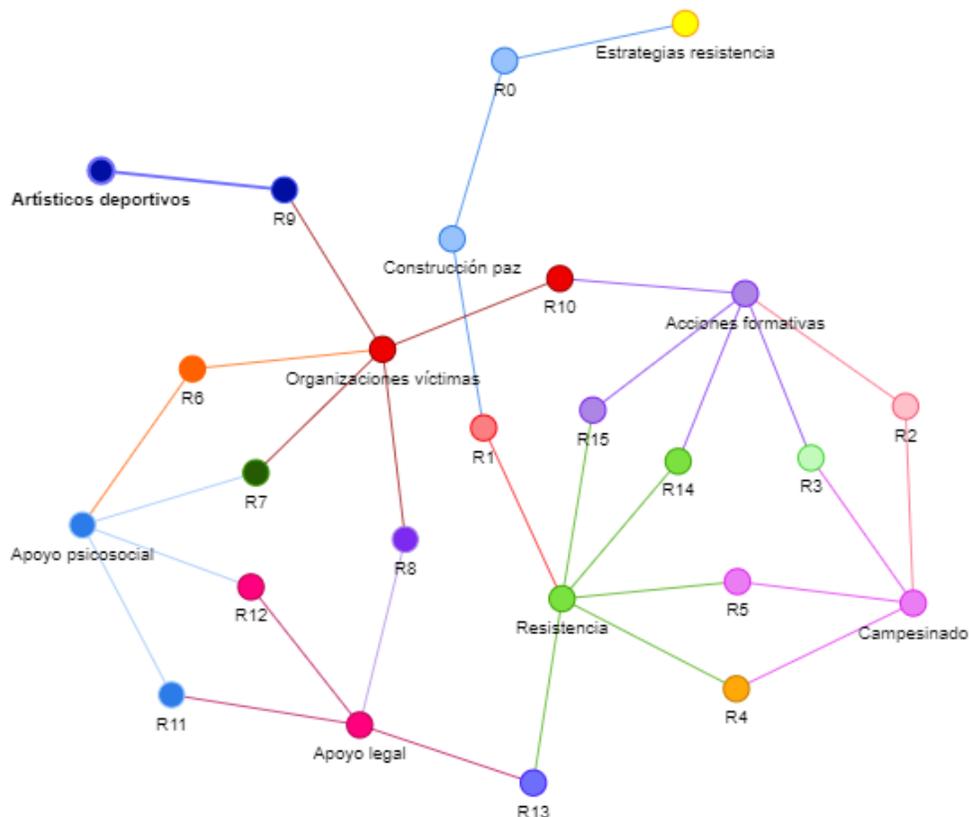


Figura 7-2: Grafo de relaciones para Resistencia $n=16$, $lift=0.6$, $confidence=0.1$

Como conclusiones generales de este informe, es importante establecer que las relaciones desarrolladas entre diferentes aspectos de las subsecciones de acuerdo con las etiquetas. Para el caso de las resistencias organizadas, imperan las relaciones y las recomendaciones surgidas a través de las organizaciones de víctimas con desarrollos unidireccionales y bidireccionales. Del mismo modo, las acciones formativas adquieren un protagonismo importante. Para el caso de las resistencias organizadas y de afrontamiento y resistencias no violentas - cotidianas, imperan las relaciones directas o consecuensales con el nodo personales. El afrontamiento en este caso se refiere a diferentes estrategias establecidas para determinar prácticas de resistencia que perduren en el tiempo y en las personas, tales como: la espiritualidad, el enfrentar violento, el silencio y la descarga emocional. Para el caso de la relación entre las formas de resistencia organizada y los hechos victimizantes, las recomendaciones fluctuaron entre diversos aspectos concernientes al conflicto armado colombiano. Entre estos se encuentran el homicidio, la amenaza al derecho a la vida, el desplazamiento forzado y las masacres. Por último, en relación con las entrevistas correspondientes a los hechos victimizantes, los afrontamientos no violentos - cotidianos y las resistencias, se encuentran varios resultados determinantes. En tanto a la regla establecida entre las tres categorías se recomienda analizar la forma en las que los afrontamientos personales y basados en las redes de apoyo funcionan en relación a las organizaciones de víctimas como formas de resistencia ante los hechos victimizantes como la amenaza verbal. En cuanto a las relaciones entre resistencia y afrontamiento, se determina en mayor medida que las formas de afrontamiento son de carácter personal y se caracterizan por ser conformadas por formas de resistencia en las que resaltan las organizaciones de víctimas y aquellas que cuentan con el apoyo psicosocial. Finalmente, en relación con las relaciones entre los afrontamientos y hechos victimizantes, se establece que las recomendaciones relacionan, en su mayoría, en los afrontamientos a nivel personal, de las redes de apoyo y el enfrentar violento; esto en su relación con los hechos victimizantes que en su mayoría son de homicidio, amenaza verbal, desplazamiento y de masacre.

Con las anteriores conclusiones se pone en evidencia la importancia de la minería de datos en los análisis de sistemas de recomendaciones y su apoyo dentro de las investigaciones relacionales. Esto se debe a que cuando se habla del manejo de datos resulta indispensable conocer el concepto de la minería de datos, la cual se entiende como el proceso de extracción de la información dentro de los grandes volúmenes de datos analizados. Londhe, Mahajan y Bhojar [6], pues es determinante para descubrir patrones y tendencias desde la gestión del conocimiento (GC) en las empresas de hoy, toda vez que resulta fundamental dados los escenarios de competitividad y desarrollo que deben enfrentar [7]. Allí, el cruce de datos pretende sacar el máximo provecho al tejer información de las bases de datos de los usuarios y sus comportamientos para analizar las distintas variables y establecer correlaciones que brinden información de conductas en ocasiones no tan evidentes [8].

Bibliografía

- [1] R. Czajka, “A revista civilização brasileira: projeto editorial e resistência cultural (1965-1968),” *Rev. Sociol. Polit.*, vol. 18, Feb. 2010 [Online]. url: <https://www.scielo.br/j/rsocp/a/h74gXbhKQtQdbkvyX8xP9XC/?lang=pt>.
- [2] Comisión de la Verdad, “El campesinado ha sido parte central de las resistencias a la guerra,” Mar. 16, 2022 [Online], <https://comisiondelaverdad.co/actualidad/noticias/presentacion-informe-guerra-contrael-campesinado>.
- [3] Ministerio de Salud, “Atención psicosocial a víctimas,” s.f. [Online], https://www.minsalud.gov.co/proteccionsocial/Paginas/Victimas_Atenc_Psicosocial.aspx.
- [4] Unidad para la integración y la reparación integral a las víctimas, “Participación de las víctimas,” s.f. [Online], <https://www.unidadvictimas.gov.co/es/participacion-de-las-victimas/283>.
- [5] ASDI, “El abc de la ley de víctimas,” *Programa de las Naciones Unidas para el desarrollo - PNUD*, vol. 7, Ago. - Sep. 2011 [Online]. url: <https://premar.ucaldas.edu.co/organizaciones-de-victimas/>.
- [6] S. R. Londhe, M. R. A., and B. J. Bhojar, “Survey on mining high utility itemset from transactional database,” *The International Journal of Innovative Research and Development*, vol. 2, no. 13, pp. 43–47, 2013.
- [7] C. E. Marulanda, M. López, and M. H. Mejía S., “Minería de datos en gestión del conocimiento de pymes de colombia,” *Revista Virtual Católica del Norte*, vol. 1, no. 50, pp. 224–237, 2017.
- [8] E. Terán, H. Elbert, M. Alcivar, and A. Puris, “Aplicaciones de minería de datos en marketing,” *Revista Publicando*, vol. 3, no. 8, pp. 503–5012, 2016.

8 Conclusión

Aunque cada capítulo cuenta con sus propias conclusiones, es posible determinar que a través del uso de diversas metodologías se pueden desarrollar análisis gracias al procesamiento de lenguaje natural dentro de las diversas entrevistas que componen los múltiples corpus de la Comisión de la Verdad. Gracias a ello, se pudieron obtener resultados de varias índoles donde sale a relucir el uso del lenguaje.

Gracias a todos los resultados proporcionados, es evidente que en todos los estudios la importancia del lenguaje es transversal e intachable, puesto que los diferentes análisis involucraron de alguna u otra manera la forma por la cual se describen los acontecimientos producidos durante el marco del conflicto armado dentro de los diferentes corpus de la Comisión. Ya sea mediante un análisis de frecuencias, impactos, promedios o métricas, los componentes del lenguaje fueron determinantes a la hora de establecer relaciones y conclusiones dentro de cada informe. No obstante, la palabra componente puede parecer bastante amplia, los factores determinantes dentro del lenguaje fueron las enunciaciones, los descriptores y las acciones acontecidas dentro de las entrevistas, puesto que, mediante el uso de estos mecanismos, se pueden rescatar los verbos, las personas, las organizaciones, los adjetivos y los lugares más recurrentes y disidentes dentro del corpus.

De este manera, se permite el espacio de generar relaciones transversales a lo largo de las entrevistas, entendiéndolas no como agentes aislados, sino como producciones de significados compartidos dentro de los diversos aspectos acarreados dentro del corpus analizado en cada caso. A su vez, se dan pasos más cerca en el rol del esclarecimiento de lo ocurrido al interior del país.

Otro aspecto interesante por concluir viene siendo la importancia de la interpretación de los diferentes análisis desarrollados, puesto que estos pueden desarrollarse en cronologías y geografías distintas, lo cual puede ayudar a establecer la presencia del conflicto armado en las regiones y periodos colombianos. Así, los análisis de las diferentes entrevistas que componen los múltiples corpus de la Comisión, podrían develar aspectos como presencia armada, actos de sevicia y demás. Es importante delimitar esto, para no incurrir en el error de establecer una generalidad concluyente en toda la región, cuando en los aspectos regionales y locales se puedan estar gestando otras dinámicas en torno al conflicto armado. Un ejemplo de esto, es la relación de lo urbano y lo rural, no se le puede dar la misma connotación del conflicto

armado de la misma manera a ambas regiones, es necesario establecer las características de diferencia.