

**ESTRATEGIA PARA EL ESTUDIO COMPARATIVO DE TEXTOS EN
CONTEXTOS POLÉMICOS. El caso de la controversia entre Sergio Elías Ortiz
y José Rafael Sañudo**

JORGE JAVIER MARTÍNEZ RUBIO

**UNIVERSIDAD DE NARIÑO
MAESTRÍA EN ETNOLITERATURA
SAN JUAN DE PASTO
2011**

**ESTRATEGIA PARA EL ESTUDIO COMPARATIVO DE TEXTOS EN
CONTEXTOS POLÉMICOS. El caso de la controversia entre Sergio Elías Ortiz
y José Rafael Sañudo**

JORGE JAVIER MARTÍNEZ RUBIO

**Trabajo de grado presentado como requisito parcial para optar al título de
Magíster en Etnoliteratura**

**Asesor:
Mag. Melba Rivas Benavides**

**UNIVERSIDAD DE NARIÑO
MAESTRÍA EN ETNOLITERATURA
SAN JUAN DE PASTO
2011**

NOTA DE RESPONSABILIDAD

Las ideas y conclusiones aportadas en el trabajo de grado son de responsabilidad exclusiva del autor.

Artículo 1° del Acuerdo 324 de octubre 11 de 1966, emanado del Honorable Consejo Directivo de la Universidad de Nariño.

Nota de aceptación

Firma del Presidente de Tesis

Firma del Jurado

Firma del Jurado

San Juan de Pasto, Octubre de 2011

AGRADECIMIENTOS

Decir que la elaboración de una tesis de maestría es un proceso intermitente en intensidad y difícil en muchos momentos no es más que constatar algo obvio para cualquiera que haya alcanzado el grado de magister. Decir que el esfuerzo que supone culminar esta tarea resulta esencial para madurar, tanto en el ámbito profesional como personal, es seguramente una reflexión que otros muchos habrán hecho en circunstancias similares. Pero el tópico que con seguridad es el más repetido, y sin lugar a dudas el más cierto cuando se refiere a trabajos que se extienden en el tiempo, es que uno jamás podría haber llegado al resultado final sin la complicidad, la paciencia y la ayuda de un amplio grupo de personas. El apoyo de todos y cada uno, expresado en una multitud de formas diferentes, ha sido fundamental para llegar hasta aquí y es no solo una obligación, sino un autentico placer poder expresar mi más sincero agradecimiento.

En primer lugar agradecer a mi directora de Tesis Mg. Melba Rivas, que me hizo indagar en el campo del estudio del texto y apoyó para que esta tesis llegara finalmente a buen puerto.

Gracias a los directivos y a los docentes del Programa de Etnoliteratura por el apoyo y la transferencia de conocimientos.

Y sobre todo, quiero agradecer el apoyo especial de mi familia, a mi esposa Yany y a mi pequeño Javis por la tolerancia y la espera.

RESUMEN

La crítica generada por los ESTUDIOS SOBRE LA VIDA DE BOLIVAR, manifiestan en su recepción y detracción, la condición del discurso como medio y como objetivo en el transcurrir en los enfrentamientos políticos. Esta situación ha convertido a las formulaciones discursivas, como la que es objeto de este trabajo, en actos potencialmente peligrosos, por ello nuestra sociedad las ha investido y cercado con múltiples cauciones y procedimientos que pretenden controlarlas y apropiarse de ellas. Estos procedimientos pueden ser externos al discurso mismo, operaciones de exclusión -como la descalificación del discurso por ser emisión de un “fanático, sectario, o insensato”¹- o de separación -como la partición de discursos en verdaderos y falsos²- o puede tratarse de intervenciones inherentes al discurso, que esquivan y domeñan el poder de irrupción, novedad y circulación de los enunciados -como la técnica del comentario, la distancia del autor o la distribución del discurso en disciplinas con fronteras reconocibles. Un programa de investigación en este campo implicaría analizar y cuestionar críticamente estos constreñimientos internos y externos, haciendo aflorar las formaciones discursivas -no uniformables, por ejemplo, bajo la etiqueta del autor o las reglas de la disciplina- como violencia y poder operado sobre las cosas y sobre los seres humanos- y no como mensaje que aproxima y permite una conciliación comunicativa- surgiendo y funcionando en relación con un campo exterior e eventos discursivos y no discursivos -no como manifestación de una interioridad, de un sujeto que se expresa a través de ellos-.

Estudiar una obra polémica, como los “Estudios sobre la vida de Bolívar”, lleva inherente el riesgo de multiplicar las querellas suscitadas, e incitar el afloramiento de simpatías y antipatías. Este escenario ha motivado el desarrollo de estrategias metodológicas con técnicas y paradigmas existentes, y la integración de los métodos de análisis que faciliten el proceso “objetivo” de exploración de datos textuales.

¹ Son sintomáticas las imágenes L. E. Nieto Caballero en “Sañudo contra Bolívar” refiriéndose a José Rafael Sañudo: *“arremete contra los infractores en forma airada y persistente, sin tomar resuello, congestionado el rostro, apretados los dientes, los ojos llenos de sangre. No se conforma con el primer ataque. No le importan los ayes de las víctimas”*. (Hay que tener en cuenta que según los comentaristas, LENC fue de aquellos que mostró más benevolencia con Sañudo). Eduardo Castillo, en su crítica “Bolívar ante la historia”, recoge los epítetos con los cuáles se tilda a Sañudo: *“fanático, insensato y sectario”*

² Sañudo en la “Advertencia de la Segunda Edición” de sus Estudios, escribía: *“Pocos días después de que publiqué esta obra, decía a un inteligente sacerdote que el 70% de mis afirmaciones quedarían en pie y que sólo un 30% sería refutado, porque juzgaba que, ora con publicar nuevos documentos, ora con raciocinios muy ajustados a la lógica y este porcentaje destruirían los escritores que hablaran de ella; ahora empero que han pasado más de seis años, puedo decir que engañéme, pues la mayor parte de esos escritores, insultos en lugar de razones prodigáronme, y los más discretos declararon su sentir pero no hicieron refutación alguna”*.

El objetivo de este trabajo es definir un marco teórico-metodológico, para el estudio de discursos que posean la condición de medio y objetivo en el transcurrir de los enfrentamientos políticos, que presente en forma sistemática la integración de las distintas técnicas estadísticas de análisis léxico y técnicas estadísticas de exploración multivariada; y utilizarlas en el trazado de un protocolo o guía para la exploración y diagnóstico de datos textuales inherentes a estos discursos. Además, experimentar con las estrategias metodológicas a un Caso Ejemplo como las réplicas y contrarréplicas entre José Rafael Sañudo y Sergio Elías Ortiz, suscitadas por la publicación en 1925 de “Estudios sobre la vida de Bolívar” que en forma de artículos divulgó la revista “Ilustración Nariñense” entre 1925 y 1928. El presente trabajo se inscribe en la línea de investigación MITO Y ETNOLITERATURA: aborda la reacción escrita de uno de los personajes eximios de nuestra región frente a un autor que se convirtió en mito, al ser considerado dentro de la categoría de los “hombres infames” por tratar de desmitologizar la historia del Libertador.

ABSTRACT

The criticism generated by the studies on Bolivar's life, manifested in their reception and detraction, the condition of discourse as a means and an aim in the passage in the political confrontation. This has become a discursive formulations, such as that is the subject of this work in potentially hazardous events, so our society has invested and surrounded with many precautions and procedures that seek to control them and own them. These procedures may be external to the discourse itself, excluding operations, such as disqualification from the issuance of speech for being a "fanatical, sectarian, or foolish" - separation or partition, as in true and false statements - or may be interventions inherent in the speech, dodging and harness the power of emergence, development and circulation of statements, as the technique of commentary, the distance distribution of the author or the discourse in disciplines with recognized borders. A research program in this field would critically analyze and question these internal and external constraints, that bring the non-uniform-discursive formations, for example, under the label of the author or the rules of discipline and violence and power-operated on things and human beings and not as a message that allows a reconciliation approaches and communicative-emerging and running in connection with an external field and non-discursive and discursive events, not as a manifestation of inner life, a subject that is expressed through them.

Studying a polemical work, such as "Studies on Bolivar's life," has an inherent risk of multiplying the complaints raised, and encourage the outcrop of likes and dislikes.

This scenario has motivated the development of technical and methodological strategies with existing paradigms, and integration of analytical methods to facilitate the process "objective" examination of textual data.

The aim of this paper is to define a methodological framework for the study of discourses that have the status of medium and objective in the passing of the political confrontation, to present the systematic integration of different statistical techniques for lexical analysis and multivariate statistical techniques of exploration, and use in the design of a protocol or guide for the examination and diagnosis of textual data inherent in these discourses. In addition, experience with the methodological strategies as a Case Example and counter-rebuttals between Jose Rafael Sanudo and Sergio Elias Ortiz, raised by the publication in 1925 of "Studies

on Bolivar's life" as articles that reported the magazine "Illustration Nariñense" between 1925 and 1928.

This work is part of the research MYTH AND ethno-literature: written response addresses one of the eminent personalities of our region from an author who became a myth, to be considered within the category of "infamous men" trying to demythologize the history of the Liberator.

CONTENIDO

	Pág.
INTRODUCCIÓN	27
1. DEFINICIÓN DEL PROBLEMA.....	31
2. ENFOQUES EN EL TERRENO DEL ANÁLISIS DE TEXTOS	37
2.1 DESCRIPCIÓN GENERAL DE LAS TECNOLOGÍAS DE ANÁLISIS DE TEXTOS.....	40
2.2 DESARROLLO DEL ANÁLISIS DE DATOS TEXTUAL	41
2.3 ANÁLISIS ESTADÍSTICO DE DATOS TEXTUALES	42
2.4 MARCO DE REFERENCIA.....	43
2.5 LA DISTRIBUCIÓN DE FRECUENCIAS Y EL CONTENIDO	44
2.6 LAS CONCORDANCIAS Y LA CONTEXTUALIZACIÓN	46
2.7 TÉCNICAS DESCRIPTIVAS DE ANÁLISIS DE DATOS MULTIVARIANTES	47
3. PROTOCOLO PARA EL ESTUDIO DE TEXTOS	50
3.1 MÉTODOS DE OBSERVACIÓN, EXPLORACIÓN Y COMUNICACIÓN APLICADOS AL ESTUDIO DE TEXTOS.....	50
3.2 ESTRATEGIAS METODOLÓGICAS PROPUESTAS	51
3.3 TIPO DE ESTUDIO	52
3.4 ESTUDIO COMPARATIVO DE TEXTOS.....	52
3.5 FASES Y ACTIVIDADES PROPUESTAS.....	52
3.6 PROPUESTA PARA EL ESTUDIO COMPARATIVO DE TEXTOS	56
3.7 DESCRIPCIÓN DEL PROCEDIMIENTO PROPUESTO.....	57
4. CASO PRÁCTICO	89
4.1 PROCEDIMIENTO.....	90
5. CONCLUSIONES	198
5.1. MEJORÍAS.....	202
5.2. RUMBOS DE INVESTIGACIÓN POSTERIOR	203
ANEXOS	213

LISTA DE TABLAS

	Pág.
Tabla 1. Artículos de la polémica entre Sergio Elías Ortiz y José Rafael Sañudo publicados en la revista "Ilustración Nariñense"	92
Tabla 2 Índice de corpus, ordenado por frecuencia absoluta (fragmento)	95
Tabla 3 Gama de frecuencia de formas (fragmento).....	97
Tabla 4 Glosario de formas sin lematización (fragmento de 4826)	99
Tabla 5 Ejemplo de lematización (fragmento)	100
Tabla 6. Gama de frecuencia de lemas.....	101
Tabla 7. Glosario de lemas seleccionados (100)	102
Tabla 8 Índice de similaridad ESTUDIOS y artículos de e revista.....	107
Tabla 9. Análisis factorial de coocurrencia simple que muestra diferencias en la utilización de vocabulario en cada uno de los artículos (fragmento).	109
Tabla 10 Análisis de correspondencia estadística para tabla cruzada por artículo	111
Tabla 11 Matriz índice de similaridad.....	117
Tabla 12 Estadísticas aglomeración documentos	117
Tabla 13 Estadísticas de similaridad del vocabulario distribuido en los dos aglomerados de documentos	118
Tabla 14 Tabla cruzada, presenta conteo de palabras por documento según el orden de su frecuencia de aparición	121

Tabla 15 Tabla cruzada tabula frecuencia total con documentos presenta porcentaje de fila agrupadas por frecuencia de palabras.....	126
Tabla 16 Tabla cruzada tabula frecuencia total con documentos presenta porcentaje de fila agrupadas por frecuencia de palabras en orden descendente hasta un chi cuadrado 20	129
Tabla 17 Presenta las palabras típicas de Sergio Elías Ortíz en el primer artículo “HISTORIA?” con el seudónimo “Un indio patriota”	130
Tabla 18. Presenta las palabras típicas de Sergio Elías Ortíz en el segundo artículo “ <i>AÚN VIVE LA HISTORIA</i> ” con el seudónimo “Alexo Corfú”	131
Tabla 19. Presenta las palabras típicas de Sergio Elías Ortíz en el tercer artículo “ <i>HISTORIA?</i> ” con el seudónimo “Un indio patriota”	132
Tabla 20. Presenta las palabras típicas de José Rafael Sañudo en el artículo “ <i>Los refutadores del señor José Rafael Sañudo</i> ” con el seudónimo “Próspero Gallo”	133
Tabla 21. Presenta las palabras típicas de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (1)</i> ” con el seudónimo “Patricio Leal y Franco”	134
Tabla 22. Presenta las palabras típicas de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (2)</i> ” con el seudónimo “Patricio Leal y Franco”	135
Tabla 23. Presenta las palabras típicas de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (3)</i> ” con el seudónimo “Patricio Leal y Franco”	136
Tabla 24. Presenta las palabras típicas de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (4)</i> ” con el seudónimo “Patricio Leal y Franco”	137
Tabla 25 Unidades léxicas exclusivas por frecuencia en artículo	138
Tabla 26 Tabla cruzada tabula presenta conteo de palabras por autor según el orden de frecuencia de aparición.	139
Tabla 27 Tabla cruzada presenta unidades lexicales típicas por autor	144

Tabla 28 Unidades léxicas exclusivas de José Rafael Sañudo ordenadas por frecuencia.....	145
Tabla 29 Unidades léxicas exclusivas de Sergio Elías Ortiz ordenadas por frecuencia.....	146
Tabla 30 proximidad lemas Bolivar e hiatoria con el resto de formas	148
Tabla 31. Frecuencia de palabras agrupadas en clusters presentes en los artículos.....	161
Tabla 32 Muestra el orden de relevancia de los términos agrupados en los cluster	163
Tabla 33 Tabla cruzada agrupaciones de palabras y artículos de Sergio Elías Ortiz y José Rafael Sañudo	164
Tabla 34 Frecuencia de las agrupaciones de palabras en el primer artículo de Sergio Elías Ortiz firmado con el seudónimo un indio patriota	166
Tabla 35 Frecuencia de las agrupaciones de palabras en el segundo artículo de Sergio Elías Ortiz firmado con el seudónimo Alexo Corfú.....	167
Tabla 36 Frecuencia de las agrupaciones de palabras en el tercer artículo de Sergio Elías Ortiz, segundo firmado con el seudónimo un indio patriota	168
Tabla 37 Frecuencia de las agrupaciones de palabras en la primera contrarréplica de José Rafael Sañudo, firmada con el seudónimo Próspero Gallo	169
Tabla 38 Frecuencia de las agrupaciones de palabras en la segunda contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco.....	170
Tabla 39 Frecuencia de las agrupaciones de palabras en la tercera contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco	171

Tabla 40 Frecuencia de las agrupaciones de palabras en la cuarta contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco	172
Tabla 41 Frecuencia de las agrupaciones de palabras en la quinta contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco	173
Tabla 42 presenta los cluster típicos con sus respectivas palabras de Sergio Elías Ortíz en el primer artículo “HISTORIA?” con el seudónimo “Un indio patriota” ...	174
Tabla 43. Presenta los cluster típicos con sus correspondientes palabras, de Sergio Elías Ortíz en el segundo artículo “ <i>AÚN VIVE LA HISTORIA</i> ” con el seudónimo “Alexo Corfú”	176
Tabla 44. Presenta las agrupaciones de palabras típicas de Sergio Elías Ortíz en el tercer artículo “ <i>HISTORIA?</i> ” con el seudónimo “Un indio patriota”	178
Tabla 45. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “ <i>Los refutadores del señor José Rafael Sañudo</i> ” con el seudónimo “Próspero Gallo”	180
Tabla 46. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (1)</i> ” con el seudónimo “Patricio Leal y Franco”	182
Tabla 47. Presenta las agrupaciones típicas palabras de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (2)</i> ” con el seudónimo “Patricio Leal y Franco”	182
Tabla 48 Presenta las agrupaciones típicas palabras de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (3)</i> ” con el seudónimo “Patricio Leal y Franco”	186
Tabla 49. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “ <i>UNAS VERDADES (4)</i> ” con el seudónimo “Patricio Leal y Franco”	188
Tabla 50. Presenta las agrupaciones de palabras según el carácter del artículo	190
Tabla 51 Asociaciones de palabras típicas según carácter de los artículos.....	191

Tabla 52 Agrupaciones típicas de palabras en los artículos de Sergio Elías Ortiz.....	191
Tabla 53 Agrupaciones típicas de palabras de José Rafael Sañudo en su contrarréplica	192

LISTA DE ILUSTRACIONES

	Pág.
Ilustración 1 Fichero de datos delimitado por ";"	93
Ilustración 2 Ejemplos de segmentos repetidos o multipalabras.....	94
Ilustración 3 Contexto de "Bolívar" y "justicia" (fragmento)	98
Ilustración 4 Gama de frecuencia de lemas	101
Ilustración 5 Fragmento de estructura léxica base (artículos*formas).....	105
Ilustración 6 Discursos extremos "Estudios sobre la vida de Bolívar" y réplicas y contrarréplicas.....	106
Ilustración 7 Distribución de la inercia de cada discurso	106
Ilustración 8 Artículos de Ilustración Nariñense: réplicas de Sergio Elías Ortiz y contrarréplicas de José Rafael Sañudo.....	107
Ilustración 9 Distribución de los discursos según el vocabulario	108
Ilustración 10 Dendograma de agrupación de los artículos estudiados	109
Ilustración 11 Gráfico de burbujas de cinco grupos por índice de similaridad en vocabulario.....	110
Ilustración 12 Mapa de artículos y vocabulario	111
Ilustración 13 Dendograma - clasificación jerárquica directa de los artículos	116
Ilustración 14 Discursos "activos" por grupo	116
Ilustración 15 Gráfico de burbujas de palabras en cada documento (fragmento).....	125
Ilustración 16 Gráfico de burbujas frecuencia de formas por autor	143
Ilustración 17. Lista de las asociaciones del lema "BOLIVAR"	147
Ilustración 18. Palabras asociadas a la forma "BOLIVAR" e "HISTORIA"	153
Ilustración 19, visualiza el árbol o Dendograma con las agregaciones efectuadas.....	155
Ilustración 20 Vocabulario agrupado y reclasificado en cada cluster.	156
Ilustración 21 Gráfico de burbujas del vocabulario,.....	157
Ilustración 22 Vocabulario por cluster	159
Ilustración 23 Presenta la distribución de los cluster de palabras en el plano.....	160

Ilustración 24. Representación del vocabulario agrupado y los artículos de Sergio
Elías Ortiz y José Rafael Sañudo..... 165

GLOSARIO

Análisis factorial de correspondencias (AC)

Se trata de un método descriptivo (no explicativo) que se clasifica entre los métodos multivariantes de interdependencia y permite visualizar los datos (que pueden ser cualitativos o cuantitativos) mediante la representación de una nube de puntos en un espacio de dimensiones reducidas, en función de las distancias geométricas entre los puntos. Técnica estadística de análisis factorial aplicada al estudio de tablas de datos cuyas celdas contienen valores de frecuencia (números reales positivos) o valores de tipo presenciaausencia ("1" o "0"). Como todos los métodos de análisis factorial, el análisis de correspondencias permite la extracción de nuevas variables "los factores" que resumen de una manera organizada la información significativa contenida en los innumerables datos de las tablas; además, esta técnica de análisis permite la creación de gráficos que muestran "en uno o más espacios" los puntos que identifican los objetos en filas y/o en columnas, que son las entidades lingüísticas (palabras, lemas, segmentos de textos y textos) con sus respectivas características de proveniencia.

Cadenas de Markov

Una cadena markoviana está constituida por una sucesión (o secuencia) de eventos, generalmente indicados como estados, caracterizada por dos propiedades: el conjunto de los eventos y de sus posibles resultados es finito; y el resultado de cada evento depende sólo (o al máximo) del evento inmediatamente anterior. Con la consecuencia de que a cada transición de un evento a otro le corresponde un valor de probabilidad.

En el ámbito de los estudios lingüísticos sus aplicaciones tienen como objeto las posibles combinaciones de las varias unidades de análisis en el eje de las relaciones sintagmáticas (una unidad tras otra). El análisis de las cadenas markovianas concierne la secuencia relativas a las relaciones entre unidades lexicales (palabras, lemas o categorías) presentes en el corpus en análisis.

Chi-cuadrado

Es un test estadístico para comprobar si los valores de frecuencia obtenidos por un examen, y registrados en una tabla cualquiera de doble entrada, son significativamente diferentes a los teóricos. El valor de umbral es 3.84 ($df = 1$; $p = 0.05$) o 6.64 ($df = 1$; $p = 0.01$).

Cluster Análisis

Los métodos de clasificación permiten la obtención de una partición en clases de las formas (u otro tipo de unidades textuales consideradas) en función de su

empleo por parte de los sujetos o la clasificación de los individuos en función del vocabulario empleado por los mismos. Involucran un conjunto de técnicas estadísticas cuyo objetivo es individuar grupos de objetos que tengan dos características complementarias: máxima homogeneidad interna (dentro de cada cluster); máxima heterogeneidad externa (entre cluster y cluster). En general, hay dos tipos de técnicas de cluster análisis: métodos jerárquicos, cuyos algoritmos reconstruyen la jerarquía completa de los objetos analizados (el llamado "árbol"), sea en orden ascendente que en orden descendente; y métodos divisorios, cuyos algoritmos proveen que el usuario haya definido previamente el número de grupos en los cuales se dividen los objetos analizados.

Cluster

A veces castellanizado como clúster, es un término inglés encontrado en varios tecnicismos. La traducción literal al castellano es "racimo" o "grupo": Clases de individuos homogéneas en cuanto al vocabulario empleado, o a la determinación de grupos de palabras que suelen ser empleadas por los mismos individuos y que delimitan, por tanto, campos semánticos o temáticas conectadas entre sí

Concordancias o Contextos Elementales

Una concordancia es, en términos generales, un sintagma (que puede coincidir con la frase o con una expansión fija de posiciones a derecha e izquierda) que contiene la ocurrencia de una o varias formas lexicales seleccionadas (y, eventualmente, cumple con uno o varios criterios suplementarios formales o de contenido). Los contextos elementales pueden ser: Frases, contextos elementales que terminan con signos de puntuación (.?!); Párrafos, contextos elementales que terminan con signos de puntuación (.?!) y retorno del carro; Fragmentos, contextos elementales de longitud comparable y compuestos de uno o más enunciados,

Corpus

Colección de uno o más textos o documentos. Un corpus de textos, es un conjunto de fragmentos de discurso escrito, cuyos componentes elementales son unidades lexicales o estadísticas - simples palabras- capaces de capturar el sentido elemental del texto. Algunos ejemplos de corpus:

- un solo texto o documento que trate cualquier tema;
- un conjunto de artículos tomados de la prensa, referentes al mismo tema;
- una o varias entrevistas realizadas en el mismo proyecto de investigación;
- un conjunto de respuestas a una pregunta abierta de un cuestionario;
- una lista de direcciones sacada de internet;
- uno o varios libros del mismo autor que afronten temas similares;
- un conjunto de respuestas a una pregunta abierta de un cuestionario;

-transcripciones de focus groups.

Datos complementario

Variable cualitativa categórica o variable cuantitativa continua que describen o aportan información accesoria para los textos. Las primeras son las etiquetas usadas para identificar y clasificar diferentes partes del corpus: nombres con características que identifican tipos de sujetos, de textos y de contextos. Cada variable tiene dos o más modalidades, cada una de las cuales, de manera inequívoca, corresponde a un valor de codificación: por ejemplo, la variable "sexo" tiene dos categorías (masculino y femenino).

Delimitadores

Se pueden distinguir dos tipos de delimitadores: fuertes y débiles. Toda sucesión de palabras no separadas por un delimitador fuerte es un segmento. Un segmento que se repite al menos dos veces, es un segmento repetido del corpus

Desambiguación

Operación que intenta resolver casos de ambigüedad semántica, concretamente los atribuibles a los homógrafos, es decir, palabras con la misma forma gráfica pero con diversos significados. Es decir que la desambiguación del sentido de las palabras es identificar el sentido correcto de una palabra en un contexto.

Especificidad

Es el nombre de un proceso que nos permite comprobar cuáles son las unidades lexicales (palabras, lemas o categorías) típicas o exclusivas de un texto, o de un subconjunto del corpus definido por una variable cualquiera. Las unidades lexicales "típicas" se definen por exceso (sobre utilización) o por defecto (sub utilización), según un criterio estadístico. Las unidades léxicas "exclusivas" son las presentes solamente dentro del subconjunto considerado y "no" en otros.

Forma gráfica o léxica

La forma gráfica es una unidad que se define sucesión de caracteres no delimitadores (en general letras) comprendidos entre dos delimitadores (blancos y signos de puntuación). Constituye una unidad de análisis frecuente utilizado, por la simplicidad de su recuento, por la claridad de su definición y, también, por la información que contiene. El conjunto de formas de un texto constituye su vocabulario.

Hapax

Una palabra o segmento se caracteriza por un número de frecuencia u ocurrencia y por las posiciones en el documento (localización). Una forma empleada una vez se llama Hapax.

Homógrafos

Dos o más palabras son homógrafas cuando tienen la misma forma gráfica (se escriben de la misma manera) pero tienen diversos significados.

Indexación

Mediante la indexación se construye el glosario, tanto por orden de frecuencia como por orden alfabético. Este se presenta en una tabla de orden lexicométrico donde se muestra el número identificador de cada palabra, la palabra del glosario del corpus, la frecuencia de aparición y la longitud de la unidad medida en número de caracteres.

Índices de Asociación

Los índices de asociación (o de similitud) se utilizan para analizar las concurrencias de las unidades lexicales en el interior de los contextos elementales, es decir datos binarios del tipo presencia/ausencia.

Individuo

Individuo estadístico, caso de estudio, texto elemental o individual, partición del corpus, documentos Primarios que corresponden a las unidades de contexto precedidas por una fila de codificación. Cada subconjunto se define por medio de una modalidad y de una variable. Algunos ejemplos: un capítulo de un libro, un artículo periodístico publicado en el mismo año; unas respuestas a una pregunta abierta.

Inercia

Se utiliza la inercia, como distancia entre individuos las mismas utilizadas en los métodos factoriales, la inercia es un índice de deformación de la nube. Se tiene son dos representaciones que contienen la información de la tabla de contingencia: la nube de perfiles fila y la nube de perfiles columna, con puntos ponderados, centradas y con una inercia asociada. En muchos casos, cada observación es un punto de un espacio euclídeo p -dimensional, se emplea en este espacio euclídeo una distancia d , para valorar la proximidad entre dos puntos puedo calcular promedios (centros de gravedad) y puedo medir la calidad de una partición mediante: la inercia intraclases que mide el parecido de los individuos

dentro de cada clase (cuanto más pequeña sea, mejor es la partición); ó la inercia interclases que mide lo diferentes que son unas clases de otras (cuanto más grande sea , mejor es la partición).

Infometría

Es la ciencia de la aplicación de los métodos matemáticos a los hechos y situaciones que se producen en el campo de la información, para describir y analizar sus fenómenos, descubrir sus leyes y servir de soporte a sus decisiones. y se inscribe en lo que se llama "Descubrimiento de Conocimientos en las Bases de Datos" que se define como "la extracción, a partir de datos, de una información implícita, desconocida, potencialmente útil".

Isotopía

Isotopía (iso = igual; topos = lugar) se refiere a un concepto de significado como "efecto del contexto", es decir, como algo que no pertenece a las palabras consideradas aisladamente, sino como resultado de sus relaciones en el interior de los textos. La función de las isotopías es la de facilitar la interpretación de los discursos o de los textos; de hecho, cada una de ellas detecta un contexto de referencia común a varias palabras, que no derive de sus significados específicos. La detección de una isotopía, por lo tanto, no es la mera observación de un "dato", sino el resultado de un proceso de interpretación

Lema y Lematización

Un lema se define como una palabra con la misma raíz lexical (el lexema) y que pertenece a la misma categoría gramatical (verbo, adjetivo, etc.). Constituye la "raíz léxica" que corresponde a una sola forma o a sus diferentes flexiones, siempre que sean portadoras de significado equivalente para las finalidades del estudio; o que corresponde con diferentes formas con igual significado La lematización exige que las formas del verbo se pongan en infinitivo, los sustantivos en singular, etcétera. La técnica de lematización nos permite mantener la misma información semántica de los textos a tratar, disminuyendo el tamaño de los documentos a procesar. Además, suponemos, que al sustituir una palabra por su lema, estamos concentrando la información semántica dándole el peso real a cada uno de los lemas que aparecen.

Lexia y Lexicalización

La lexía es una expresión constituida por una o más palabras que se comportan como una unidad lexical con significado autónomo. Los tipos fundamentales son tres: simple, correspondiente a la palabra en el sentido común del término (ej. "caballo", "comía"); compuesta, constituida por dos o más palabras integradas en una única forma (ej. "biotecnologías", "videoregistrador"); compleja, constituida

por una secuencia en vía de lexicalización (es. "a mi juicio", "complejo industrial"). La lexicalización es el proceso lingüístico a través del cual un sintagma o un grupo de palabras se convierten en una sola unidad lexical.

Lexicometría o Estadística textual

Están apoyados en las técnicas estadísticas desarrolladas por la escuela francesa de análisis de datos (analyse des données), se refiere a procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y operar algún tipo de análisis estadístico a partir de los resultados de tales recuentos. Se recurre a la cuantificación de los textos desde el primer momento, sin que medien operaciones de codificación previas

Núcleos Temáticos

Pequeños clusters de palabras, co-ocurrentes en los contextos elementales del corpus, que - en los mapas - se representan con las respectivas "cabezas de serie", o sea con los elementos (sustantivos o verbos) con el valor más alto de ocurrencia.

Ocurrencias y Co-ocurrencias

Las ocurrencias, son las cantidades que resultan del cómputo de cuántas veces (frecuencias) cada unidad lexical se repite dentro del corpus o dentro las unidades de contexto que lo constituyen. Las co-ocurrencias son las cantidades que resultan del cómputo del número de veces que dos o más unidades lexicales están presentes contemporáneamente en los mismos contextos elementales

Palabras clave

Son Palabras Clave todas las unidades lexicales (palabras, lemas, lexías, categorías) que, cada vez, se incluyen en las tablas a analizar.

Palabras funcionales,

Artículos, preposiciones, Conjunciones, demostrativos, algunos adverbios, etc
Palabras vacías

Muchas palabras se definen "vacías" porque solas no tienen ningún contenido específico y/o significativo. No existe un criterio estándar para construir una lista de estas palabras, algunos ejemplos que se podrían tomar: adjetivos indefinidos; artículos; adverbios; exclamaciones; interjecciones; preposiciones; pronombres demostrativos, indefinidos y relativos); verbos auxiliares (ser, haber); verbos modales (deber, poder, saber, soler, querer), etc

Palabras y Lemas

La denominada “palabra”, contiene la transcripción de las unidades lexicales o formas (palabras individuales, lexias o multi-palabras) como “cadenas” reconocidas. La denominada “lema”, contiene las etiquetas con las que están reagrupadas y clasificadas las unidades lexicales. Según los casos, un lema puede ser: el resultado del proceso de lematización automática; una voz de un “diccionario personalizado”; una categoría que indica un grupo de sinónimos; una categoría de análisis del contenido; etc.

Procesamiento de textos

El procesamiento de textos considera una gran diversidad de tareas, desde muy simples, como la separación de palabras, hasta muy complejas como algunas tareas de minería de texto. Incluye procesos de preparación de textos, búsqueda y extracción de información y descubrimiento de conocimiento.

Tabla de dato

Las tablas de datos (o matrices) se componen de filas, de columnas y de los valores registrados en las celdas respectivas permiten sintetizar - de una manera ordenada - tanto las observaciones que hay que someter a análisis estadísticos (input), como los resultados obtenidos por su aplicación (output). Según los tipos de análisis, las tablas pueden ser de tres tipos, correspondientes a otras tantas maneras de construir cruces entre filas y columnas:

- formas en fila y textos (o variables) en columna;
- textos (o fragmentos de textos) en fila y formas en columna;
- formas tanto en fila como en columna.

Tabla léxica agregada.

Cuando cada columna de esta tabla se construye a partir del agrupamiento de varios textos elementales, por ejemplo, todas las respuestas de los individuos de una determinada edad. Permite al investigador, siempre que los textos tengan una extensión similar, realizar comparaciones entre las frecuencias alcanzadas en cada uno de ellos por determinadas formas relevantes para el estudio.

Tablas léxicas

Formas de reorganizar la información que presentan interés de cara a la descripción. Una tabla léxica es una tabla de doble entrada en la que las filas (o columnas) corresponden a las unidades resultantes de la Segmentación del texto,

generalmente palabras, y las columnas (o filas) a los textos Elementales considerados.

Umbral de frecuencia

Se calcula un umbral mínimo de frecuencia para seleccionar las palabras (o los lemas) que serán introducidas en los análisis.

Unidad de Análisis

Las unidades de análisis son de dos tipos: unidades lexicales y unidades de contexto. Las unidades lexicales son formas gráficas, simples o "múltiple", archivadas y clasificadas en base a algún criterio. Las unidades de contexto son porciones de texto en las que se puede dividir el corpus; pueden ser de tres tipos: 1 documentos primarios correspondientes a la subdivisión "natural" del corpus (ej. entrevistas, artículos, respuestas a preguntas abiertas, etc.), o sea a los contextos iniciales definidos por el usuario; 2 contextos elementales, correspondientes a unidades sintagmáticas de una o más frases y definidas de modo automático (o semi-automático).

Por tanto, cada documento primario está constituido por uno o más contextos elementales;

3 subconjuntos del corpus que corresponden a grupos de documentos primarios atribuibles a la misma "categoría" (es. entrevistas de "hombres" o de "mujeres", artículos de un determinado año o de un determinado periódico, y así sucesivamente).

Unidades verbales básicas

Unidad básica de análisis, unidad física, fácilmente reconocida por un ordenador y que propicia, por tanto, una segmentación automática. Es la unidad más frecuentemente empleada y generalmente coincide con una palabra.

Valor Test

Ésta es una medida estadística utilizada para facilitar la interpretación de las polaridades factoriales detectadas con el análisis de correspondencias. Y la validación de la caracterización de grupos y cluster. Esta medida tiene dos propiedades significativas: un valor umbral (1.96), correspondiente al nivel estadístico usado más comunemente ($p = 0.05$), y un signo (-/+).

Variable léxica

Los métodos de análisis estadísticos multidimensionales complementan las antiguas técnicas de lexicometría, realizando el tratamiento de los textos considerando una nueva variable léxica, cuyas distintas modalidades serán las formas léxicas o más exactamente, las formas gráficas del corpus tratado

Vocabulario del corpus

Todo el conjunto de palabras de un corpus constituye el vocabulario del corpus. El número de ocurrencias de un corpus es la longitud del mismo.

Zipf

La frecuencia de palabras se asocia generalmente al nombre de Zipf por haber establecido una ley fundamental que afirma que, en cualquier texto, el producto de la frecuencia de cada palabra por su rango es constante.

INTRODUCCIÓN

Foucault ha develado el hecho de que en toda sociedad la producción del discurso está a la vez controlada, seleccionada y redistribuida por un cierto número de procedimientos que tienen por función conjurar los poderes y peligros, dominar el acontecimiento aleatorio y esquivar su pesada y temible materialidad.

En una sociedad como la nuestra son bien conocidos los procedimientos de exclusión. El más evidente, y el más familiar también, es lo prohibido. Se sabe que no se tiene derecho a decirlo todo, que no se puede hablar de todo en cualquier circunstancia, que cualquiera, en fin no puede hablar de cualquier cosa. Tabú del objeto, ritual de la circunstancia, derecho exclusivo o privilegiado del sujeto que habla: he ahí el juego de tres tipos de prohibiciones que se cruzan, se refuerzan o se compensan, formando una compleja malla que no cesa de modificarse.

Una de las regiones en las que la malla está más apretada, en la que se multiplican los compartimentos negros, son las de la política: como si el discurso, lejos de ser ese elemento transparente o neutro en el la política se pacifica fuese más bien uno de esos lugares en que se ejercen, de manera privilegiada, algunos de sus más temibles poderes. El discurso, por más que en apariencia sea poca cosa, las prohibiciones que recaen sobre él, revelan muy pronto, rápidamente, su vinculación con el deseo y con el poder. Y esto no tiene nada de extraño: ya que el discurso -el psicoanálisis nos lo ha mostrado- no es simplemente lo que manifiesta (o encubre) el deseo; es también lo que es el objeto del deseo; y ya que -esto la historia no cesa de enseñárnoslo- el discurso no es simplemente aquello que traduce las luchas o los sistemas de dominación, sino aquello por lo que, y por medio de lo cual se lucha, aquel poder del que quiere uno adueñarse.

Precisamente el estudio de los discursos podría ofrecer artilugios metodológicos para el análisis de esa interrelación entre saber, deseo y poder.

Textolingüísticamente todo discurso expresa un sistema de “mundos lexicales” que organiza una racionalidad y da coherencia a todo lo que el locutor enuncia. El término “mundo lexical” es una noción primaria o precategorial que remite a la concatenación de las palabras que componen un discurso determinado. Un mundo lexical es evocado por el conjunto de palabras que constituyen una frase o un fragmento del discurso, independientemente de su construcción sintáctica. (Alba, 2004).

Se propone observar la noción de mundo lexical a partir del análisis estadístico de los usos discursivos, o más precisamente, de los usos del vocabulario en uno o más textos. Los mundos lexicales pueden estudiarse entonces a través del análisis de la organización y distribución de las palabras principales co-ocurrentes en los enunciados simples de un texto. Es decir, que la estrategia metodológica se

focaliza en la distribución estadística de sucesiones de palabras que componen los enunciados de un texto, sin tomar en cuenta la sintaxis del discurso, sino únicamente la “co-ocurrencia” o presencia simultánea de varias palabras funcionales o principales (Reinert, 1993) (sustantivos, adjetivos, verbos) en un mismo enunciado, eliminando del análisis las palabras relacionales (conjunciones, preposiciones, artículos, etc.)³.

Cabe señalar que el análisis de las palabras principales co-ocurrentes en los enunciados de un texto no equivale a un análisis temático del discurso, no toma en cuenta la construcción sintáctica de la frase, sino únicamente la copresencia de las palabras funcionales en fragmentos de un texto o “unidades de contexto”⁴. Las unidades de contexto no necesariamente coinciden con la frase, sino que pueden estar constituidas por una sucesión de palabras principales contenidas en varios enunciados. Para Reinert es la lista de palabras principales lo que constituye la huella de los mundos lexicales del discurso, mientras que el contenido temático depende de la organización sintáctica y semántica de los enunciados.

Las estrategias metodológicas, en este caso, parten del supuesto de que el análisis de las sucesiones de palabras principales en un conjunto de enunciados permitirá diferenciar globalmente los “lugares de enunciación” o mundos lexicales más significativos del discurso. Algunos mundos lexicales son más evocados que otros, y para observarlos habrá que analizar la frecuencia de aparición de los conjuntos de palabras principales asociados entre sí que componen un texto.

La idea es que al utilizar un vocabulario determinado el locutor convoca un “lugar” de enunciación, el cual se define por oposición a otros lugares; de suerte que un mundo lexical no se define en sí mismo, sino en relación con otros. La oposición entre los mundos lexicales de los enunciadores se inscribe en una discriminación del vocabulario que se realiza estadísticamente.

Técnicamente los mundos lexicales son un conjunto de palabras principales que tienen una organización habitual (repetitiva) en el discurso y que se refieren a algo similar. El objetivo es precisamente poner en evidencia los mundos lexicales más frecuentes que componen un discurso. “No se trata de comparar las distribuciones estadísticas de las palabras en diferentes textos, sino de estudiar la estructura formal de sus co-ocurrencias en los enunciados de un texto dado” (Reinert, 1993, pág. 9).

³ REINERT. Justifica la eliminación de este último tipo de palabras apoyándose en el supuesto de que las palabras principales “son más aptas para expresar nuestros usos del mundo”, mientras que las palabras relacionales juegan un papel secundario para ello. 1993.

⁴ Ibíd. Ha preferido utilizar una heurística estadística para delimitar no enunciados sino “unidades de contexto elementales”, que son segmentos de texto compuestos por sucesiones de palabras principales.

Los problemas a abordar pueden surgir del estudio de textos (comparación de estilos, atribución de autor, búsqueda documental, etc.) o ser de naturaleza no textual, pero cuyo tratamiento lleve a considerar ciertos textos como datos portadores de información. Este tipo de metodologías, tiene sus orígenes en la aplicación de técnicas estadísticas al análisis de textos, discursos, análisis de contenidos, entre otros. En general, a las técnicas de Análisis Cualitativo, utilizadas en las áreas de Ciencias Sociales desde hace varias décadas.

En este trabajo la atención metodológica se centra en diversos procedimientos y técnicas lexicométricos y estadísticos utilizados en el análisis textual.

El Análisis Estadístico de Datos Textuales⁵ se refiere a procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y operar algún tipo de análisis estadístico a partir de los resultados de tales recuentos. Se recurre a la cuantificación de los textos desde el primer momento, sin que medien operaciones de codificación previas.

EL ANÁLISIS ESTADÍSTICO DE DATOS TEXTUALES comporta una serie de herramientas que se enmarcan en el análisis estadístico multidimensional descriptivo, frecuentemente llamado “Análisis de datos”. El enfoque de estas herramientas no lleva a emitir aserciones apoyadas en pruebas estadísticas sino a subrayar diferentes rasgos presentes en las observaciones que permiten orientar investigaciones posteriores y/o emitir nuevas hipótesis. Los resultados así obtenidos presentan una gran riqueza y diversidad.

Los métodos presentados en este trabajo facilitan la exploración, la gestión y la descripción de corpus permiten derivar información de ellos desde un punto de vista estadístico y poner de manifiesto sus rasgos estructurales.

Los objetivos del trabajo de tesis son:

Definir un marco teórico-metodológico, para el estudio de discursos que posean la condición de medio y objetivo en el transcurrir de los enfrentamientos políticos, que presente en forma sistemática la integración de las distintas técnicas estadísticas de análisis léxico y técnicas estadísticas de exploración multivariada; y utilizarlas en el trazado de un protocolo o guía para la exploración y diagnóstico de datos textuales inherentes a estos discursos.

En este texto además de estudiar los métodos conocidos tradicionalmente bajo el nombre de estadística textual, se hace una introducción a otra técnica complementaria para el tratamiento de información textual como es el método de

⁵ Los enfoques lexicométricos o de la estadística textual están apoyados en las técnicas estadísticas desarrolladas por la Escuela Francesa de Análisis de Datos (Analyse des Données), (Benzécri, J. P. 1973, 1976).

las palabras asociadas. La propuesta en este trabajo es presentar de forma sistemática, los métodos lexicométricos y de estadística multivariada utilizada en el análisis de textos, con el objetivo de seleccionar un conjunto de métodos complementarios cuyo ensamblaje constituya una herramienta de análisis comparativo de textos. Proponer una guía para el *enfoque de estudio de textos*, que ofrezca una *aproximación comparativa*: que permita describir, analizar e interpretar las diferencias entre los textos, con base en los principios de la estadística textual.

Transformar lo cuantificable, del texto, en significativo y volver al contenido al final del análisis para ofrecer la posibilidad de explorar el “texto bruto”, evitando que las preconcepciones del investigador “contaminen” el análisis (sobre todo en contextos polémicos) y manteniendo una condición de transparencia del contenido respecto al análisis que permita descifrar la *información sobre el significado de las palabras*, según los objetos observados, contenida en el propio documento y reintroduciendo, en el momento de *interpretar los resultados*, lo que se conoce de la lengua, de las palabras y de los propios textos.

Experimentar con las estrategias metodológicas propuestas a un Caso Ejemplo: las réplicas y contrarréplicas entre José Rafael Sañudo y Sergio Elías Ortiz, suscitadas por la publicación en 1925 de “Estudios sobre la vida de Bolívar” que en forma de artículos divulgó la revista “Ilustración Nariñense” entre 1925 y 1928. La propuesta en este trabajo se centra en el estudio comparativo de textos.

El contenido de la tesis presenta la siguiente distribución:

Capítulo 1: Plantea los principales problemas y soluciones propuestas en relación con el estudio del discurso.

Capítulo 2: Se realiza una descripción de la situación (trabajo, operación) que requiere el uso de un marco metodológico para el análisis de datos textuales. Descripción conceptual de la lexicometría

Capítulo 3: Se centra en explicar las estrategias metodológicas para el análisis de textos, en primer lugar se explica el tipo de textos según el objetivo de estudio que determinan el enfoque de “estudio comparativo de textos” a seguir, luego se detallan los procedimientos y actividades implicadas en este enfoque y por último se detallan los métodos léxicos e infométrico utilizados en este “protocolo” para el aprovechamiento de datos textuales.

Capítulo 4: Centra su atención en el caso práctico para experimentar con las estrategias metodológicas propuestas.

Capitulo 5: Se presentan las conclusiones y líneas de investigación futuras.

1. DEFINICIÓN DEL PROBLEMA

Los ESTUDIOS SOBRE LA VIDA DE BOLIVAR, vehiculan una serie de situaciones inherentes a los conceptos y procedimientos de la historiografía, contemporánea al autor de ese escrito, a sus detractores, y ahora a nosotros.

En primer lugar, se cuestiona que la Historia se identifique con un relato lineal de acontecimientos⁶. El tiempo experimentado por la historiografía actual no es lineal ni homogéneo, se reconocen múltiples ritmos y tipos distintos de transformación.

La discontinuidad no es hoy el obstáculo que el investigador debe reducir y superar, sino el instrumento y el objeto de análisis que permite diferenciar series, niveles de duración, modos de ruptura y desplazamiento. Al alejarse de la sucesión lineal, la Historia pierde cada vez más su forma de relato; su espacio es el de la simultaneidad tanto como el de la diacronía, por eso puede desmarcarse de una intención puramente ideográfica.

En segundo lugar, se cuestiona la primacía de episodios políticos y el papel protagonista del individuo en el proceso histórico⁷. No se sustituye la preeminencia de lo político por una determinación causal de la infraestructura económica, sino que se escinden niveles múltiples de determinación cuyas correlaciones particulares no obedecen a una jerarquía preestablecida: las coacciones y posibilidades abiertas por el medio geográfico, el peso de las fluctuaciones demográficas y los ciclos económicos, los límites que impone el lenguaje al desarrollo de las mentalidades y el pensamiento científico, etc.

En esas correlaciones entre jalones diversos, el individuo no se encuentra como fundamento (genio, héroe) sino como función derivada. Por eso mismo los historiadores dejan de leer los documentos descifrando en ellos las intenciones y propósitos de sus autores; los reparten en series, los someten eventualmente a análisis estadístico, fijan sus regularidades, establecen en ellos organizaciones, los conjuntos consistentes y las modificaciones conceptuales pertinentes. En este conjunto de discursos, la Historia parece funcionar del mismo modo que la etnología levistraussiana; se pone en entredicho la noción de hombre como fundamento empírico-trascendental y las ciencias humanas que descansan sobre esta noción.

Los exégetas de la sospecha, Nietzsche, Freud, Marx, Foucault, habían mostrado la presencia de un pensamiento, de un discurso anónimo (en las articulaciones de

⁶ La teoría de la expiación que propone José Rafael Sañudo, pretendería mantener una linealidad en los acontecimientos.

⁷ Recuérdese la afirmación de Sañudo sobre la falta de necesidad de la campaña libertadora, puesto que esta se lograría cuando estuviésemos maduros para la independencia.

la voluntad de saber, del deseo, de las relaciones de producción, del poder) que recusaba como ilusoria la identidad de la conciencia consigo misma y que daba cuenta de sus representaciones. Ello es una invitación a poner entre paréntesis la validez de todo un arsenal de categorías que los historiadores de las ideas utilizan como si se trataran de obviedades incuestionables. Estas categorías garantizan la continuidad del discurso más allá de la diversidad de sus ocurrencias y aseguran el papel fundador que respecto a él desempeña la conciencia constituyente. Así, por ejemplo, las nociones de “tradición”, “influencia”, “desarrollo-evolución”, permiten asegurar vínculos de contenido entre emisiones discursivas distantes; las categorías de “libro”, “género”, “obra” y “autor” consolidan la agrupación de preferencias discursivas en unidades homogéneas de sentido; los temas de “origen” y de la “interpretación”, que implican que lo dicho con anterioridad de un modo implícito y embrionario, permite establecer una infinita continuidad latente del discurso, más allá de todo comienzo aparente.

Al poner a un lado estas categorías que encajonan el discurso en unidades y continuidades preestablecidas, ¿qué queda?: una población de acontecimientos discursivos.

Se trata, pues, de investigar el discurso en tanto que acontecimiento, como población de enunciados. Foucault, señala que el nivel de acontecimiento, que es justamente lo que se quiere especificar con el enunciado, distingue a éste de la frase tematizada por los lingüistas y de la proposición estudiada por el análisis lógico.

La función enunciativa remite aun campo asociado para poder especificarse. En él pueden distinguirse tres espacios diferentes:

- ✓ En tanto función, el enunciado existe en relación con un espacio adyacente, compuesto por otros enunciados con los que mantiene relaciones de simultaneidad o sucesión. Son los enunciados que forman parte de una misma multiplicidad enunciativa lo que Foucault denomina formación discursiva. Sin embargo, una formación discursiva no es necesariamente una disciplina; lo principal es que la formación discursiva es siempre un sistema de diferencias, esto es, no reduce a unidad, ni homogeneiza la singularidad de cada enunciado, sino que lo especifica. No es un bloque uniforme sino un régimen de distancias entre los enunciados, un espacio de distribución, de dispersión.
- ✓ Por otra parte, la función enunciativa se ejerce, no solo en relación con un campo adyacente de enunciados, sino en conexión con un espacio correlativo que también permite articular las diferentes formaciones discursivas. Este espacio no se compone del referente u objeto común al que remiten los distintos enunciados, ni de una familia de conceptos comunes que permite agruparlos arquitectónicamente; tampoco se confunde con un sujeto unificador que serviría de soporte al campo enunciativo, y no puede identificarse con un

conjunto de temas u opciones teóricas (por ejemplo, del tipo mecanicismo vs. Vitalismo) que darían cuenta de la totalidad del campo enunciativo. ¿De qué se trata?

En primer lugar, no se trata de un referente compartido; son las condiciones (reglas de formación), no exteriores ni implícitas, sino inherentes al mismo discurso, que hacen posible hablar de una multiplicidad de objetos en el interior de una formación discursiva⁸.

En segundo lugar, este espacio correlativo está ocupado no por un edificio conceptual, sino por las condiciones (reglas de formación) que permiten el funcionamiento de múltiples formas conceptuales en el interior de las formaciones discursivas de una época⁹.

En tercer lugar, no hay un sujeto alojado en este espacio correlativo que actúe como fundamento y síntesis originaria de las diferentes preferencias enunciativas. Lo que hay es un conjunto de relaciones (reglas de formación), insertas en la misma superficie del discurso, que establecen la asignación de múltiples funciones para sujetos posibles que puedan intervenir en el discurso, lo que Foucault denomina modalidades enunciativas¹⁰.

En cuarto lugar, las distintas opciones teóricas que pueden diferenciarse en una formación (por ejemplo, utilitaristas vs. fisiócratas en el análisis de riquezas, o positivismo vs. fenomenología en el análisis de la finitud que caracteriza a la filosofía contemporánea) no desempeñan el papel de fundamento al campo discursivo. Se tratan de puntos de elección o difracción derivados y posibilitados por el conjunto de las relaciones discursivas (reglas de formación); éstas indican los desarrollos conceptuales que propician disensos teóricos y diversas formas de

⁸ Así por ejemplo, la serie de transformaciones que afectaron al espacio de la representación en la clausura de la época clásica, hizo posible la aparición de una nueva constelación de objetos en las formaciones discursivas emergentes: los organismos vivientes, el trabajo productivo, los sistemas flexivos de las distintas lenguas.

⁹ Los cuatro segmentos de la atribución, articulación, designación y derivación operaban como condiciones e existencia para un variado régimen de conceptos en el saber clásico: el verbo, el nombre, las raíces, la moneda, la distribución de bienes, la necesidad, el carácter, la estructura visible, las catástrofes de la naturaleza, etc.

¹⁰ Valgan como ejemplo los distintos niveles de especialización descritos en *Nacimiento de la Clínica*; la serie de transformaciones que afectó a los vínculos entre estos diferentes espacios, entre fines del siglo XVIII y primer cuarto del siglo XIX, hizo posible la emergencia de la miríada de funciones de sujeto que caracterizaron a la mirada clínica: “el interrogador soberano y directo, el ojo que mira, el dedo que toca, el órgano de desciframiento de los signos, el punto de integración de descripciones ya hechas, el técnico de laboratorio”

apropiarse del discurso haciéndolo intervenir en las luchas políticas y en los conflictos de intereses¹¹

- ✓ El discurso es entendido como un tipo de acción, una práctica discursiva. Las relaciones que lo individualizan como formación discursiva (reglas de formación de objetos, funciones de sujeto, conceptos, opciones teóricas) implican la intervención de un medio institucional, un conjunto de prácticas no discursivas. Este medio constituye una tercera superficie donde se ejerce la función enunciativa, un espacio complementario de la misma.

Al conjunto de existencia de los enunciados en la intersección de los tres espacios mencionados (adyacente, correlativo, complementario) y en los límites de una época formaría, utilizando la expresión de Foucault, el a priori histórico. Son las reglas que caracterizan al conjunto de las formaciones discursivas en un intervalo histórico determinado, el sistema de dispersión y de transformación de sus enunciados¹².

El saber no equivale a las ciencias o a los conocimientos de un período; es el conjunto de condiciones discursivas en el que pueden cristalizar o del que pueden desgajarse eventualmente positividades¹³, figuras epistemológicas¹⁴, ciencias¹⁵ y disciplinas formalizadas¹⁶.

¿A qué obedece esta urgencia en ceñir la dimensión del discurso como acontecimiento? ¿a qué responde esta exigencia de liberarlo respecto a las sujeciones antropológicas del humanismo?. La respuesta es de carácter político. Se podría considerar, en filiación heideggeriana, que el discurso humanista se relaciona con la voluntad de poder empeñada en excluir y eliminar todo aquello que se desvía y no se doblega a las exigencias de un modelo universalista y normativo de la humanidad. En esta mención se enfatiza el hecho de que las

¹¹ Así por ejemplo, la confrontación de intereses entre la clase terrateniente y la burguesía mercantil no sería la razón explicativa de la disparidad entre fisiócratas y utilitaristas en el interior del análisis de las riquezas; por el contrario es la serie de reglas de formación que intervienen en este conjunto discursivo lo que hizo posible esta bifurcación de ambas opciones teóricas siguiendo los lugares de una eventual apropiación estratégica por parte de los grupos sociales en conflicto.

¹² A diferencia de cualquier a priori formal, el a priori histórico es contingente, limitado, datado, y funciona en el interior del campo discursivo; no es su fundamento trascendental.

¹³ Cuando una formación discursiva se individualiza según un único sistema de reglas.

¹⁴ Cuando un sistema de enunciados se recorta y funciona como norma o criterio de validez de los restantes.

¹⁵ Cuando la figura epistemológica no obedece sólo a reglas de formación, sino a leyes propias en la construcción de sus proposiciones.

¹⁶ Cuando el discurso científico puede expresarse en un lenguaje formal.

condiciones de posibilidad de los discursos se encuentran fuera el discurso mismo; las reglas de formación de los enunciados se configuran a partir de los múltiples y cambiantes modos de ejercer el poder.

Foucault en su reflexión sobre el poder distingue entre dos paradigmas o formas de entender el poder. En primer lugar aborda la idea jurídica del poder, por la cual se atribuye al soberano el derecho legítimo que ejerce sobre los sujetos bajo la forma de contrato. En tanto que la segunda forma de entender el poder es denominada por Foucault forma disciplinaria del poder, en abierta oposición con la primera puesto que se caracteriza por ser anti soberana y anti jurídica. Se trata de una forma de control que ejerce la fuerza normalizando y creando las condiciones de vigilancia para imponer la docilidad de los sujetos.

La forma disciplinaria del poder es una forma de poder que, básicamente, no actúa sobre los sujetos, sino que los encausa hacia un horizonte de acción: No disciplina, sino que normaliza (Mendieta, 2007, pág. 140). No funciona con base en los ordenamientos jurídicos o los derechos, sino en normas y estándares que aluden a una tecnología social. Es un poder que surge con el desarrollo de las ciencias humanas y, en particular, con el de las ciencias de la normalización. De tal modo que el poder no está centrado, sino que es difuso, no es propiedad exclusiva de nadie, sino anónimo, no se ejerce sino que se trasmite y se vive (Mendieta, 2007, págs. 141-142). No se trata de pensar los espacios de poder sino las diversas narrativas que justifican dichos espacios y la relación de los sujetos con los mismos. Una tecnología política implica el despliegue de determinadas técnicas de intervención con el propósito de modificar la vida, tanto de individuos como de poblaciones. Es decir, es una forma aplicada de conocimientos, saberes y prácticas de intervención, dirigidas a cambiar la conducta de los individuos con el objetivo de tener un control regulado de los mismos al interior de la sociedad.

En este escenario, se puede comprender la condición del discurso como medio y como objetivo en el transcurrir en los enfrentamientos políticos. Esta situación ha convertido a las formulaciones discursivas, como la que es objeto de este trabajo: ESTUDIO SOBRE LA VIDA DE BOLÍVAR, en actos potencialmente peligrosos, por ello nuestra sociedad las ha investido y cercado con múltiples cauciones y procedimientos que pretenden controlarlas y apropiarse de ellas.

El problema que ha orientado este trabajo podría formularse:

¿Cómo estructurar un marco teórico-metodológico, para el estudio de discursos que posean la condición de medio y objetivo en el transcurrir de los enfrentamientos políticos, que presente en forma sistemática la integración de las distintas técnicas objetivas (estadísticas de análisis léxico y técnicas estadísticas de exploración multivariada, las cuales permitirían evitar ser juez y parte en la polémica); y utilizarlas en el trazado de un protocolo o guía para la exploración y diagnóstico de datos textuales inherentes a estos discursos?

Esta problemática genera la necesidad de desarrollar metodologías con técnicas y paradigmas existentes, y la integración de métodos de análisis que faciliten el proceso de exploración de discursos políticos polémicos, que permitan mostrar en qué consisten las diferencias, como es posible que unos seres humanos, en el interior de una misma práctica discursiva, hablen de objetos diferentes, tengan opiniones opuestas, hagan elecciones contradictorias; se trata de mostrar en que se distinguen unas prácticas discursivas de otras, de definir las posiciones y funciones que el sujeto puede ocupar en la diversidad de los discursos.

Se requiere revisar métodos de estudio de discurso, acordes a las condiciones señaladas: que se alejen de la sucesión lineal del relato; que dejando de lado las intenciones y propósitos de sus autores, permitan repartir los documentos en series, someterlos a análisis estadístico, fijar sus regularidades, establecer en ellos las organizaciones, los conjuntos consistentes y las modificaciones conceptuales pertinentes.

2. ENFOQUES EN EL TERRENO DEL ANÁLISIS DE TEXTOS

A continuación se presentan brevemente las principales tendencias metodológicas utilizadas para el análisis de texto. Se distinguen aquí cuatro grandes enfoques en el terreno del *análisis de textos*. (1) El *Análisis de Discurso*, (2) el *Análisis de Contenido*, (3) el *Análisis Estadístico de Datos Textuales* y (4) el *Análisis de Datos Cualitativos*. (Armony, 2002).

Se describen enseguida cada uno de ellos:

(1) El *Análisis de Discurso* constituye un campo en el que convergen una gran diversidad de perspectivas teóricas y disciplinarias que *indagan la forma de lo dicho*, sus géneros y sus procedimientos: la argumentación, la narración, la enunciación, etc. Sus fronteras son inciertas, a tal punto que prácticamente cualquier estudio textual puede pretender ser incluido en él.

Este campo tiene habitualmente una fuerte *impronta lingüística* y, aunque es extremadamente difícil de caracterizar de manera global, puede decirse que uno de los rasgos salientes es su marcado interés por los mecanismos de "producción de sentido". Lo que se indaga no es tanto el "qué" se dice, sino más bien el "cómo" se lo dice. Se recurre así a múltiples herramientas que pueden provenir de horizontes tan variados como la crítica literaria, la pragmática o la sociolingüística.

El *Análisis de Discurso* busca *develar en los hechos de lenguaje sus anclajes espaciales, temporales y sociales*. La fuerza de este enfoque radica sin duda en sus complejas elaboraciones conceptuales y en la fineza de sus interpretaciones. Sin embargo se apoya en dispositivos fuertemente ligados al punto de vista del investigador – su visión del lenguaje, sus presupuestos teóricos, sus criterios vagos de demostración empírica –, lo cual da lugar a *resultados demasiado dependientes del sistema de observación* y, por lo tanto, difíciles de poner en relación con resultados de otras investigaciones.

Los análisis discursivos tienden a reducir drásticamente la distancia entre los momentos de *la descripción* y el de *la interpretación*, espacio en el que normalmente debería desplegarse el esfuerzo de formalización metodológica.

(2) El *Análisis de Contenido* apunta esencialmente a la reducción de la complejidad de los mensajes, a través de técnicas de codificación, y a su objetivación mediante cálculos de distribución de frecuencias. Se trata de un enfoque que surge en el contexto de la ciencia social norteamericana de los años cuarenta, lo cual nos da una idea de la distancia que lo separa de la tradición francesa.

El *análisis de contenido* se propone acceder directamente a las *significaciones de diferentes segmentos que componen el texto*. Es una técnica de investigación para la descripción objetiva, sistemática y cuantitativa del contenido manifiesto en la comunicación.

Opera en dos fases: se empieza por *construir un conjunto de clases de equivalencia*, de temas y se examinan luego las *ocurrencias de los textos que serán sucesivamente analizados*. En una segunda fase se hacen los *conteos para cada uno de los temas previstos*.

Las unidades en un análisis de contenido pueden ser *los temas, las palabras o elementos de sintaxis o semántica*. Las unidades de descomposición para las medidas cuantitativas variaran también: *palabra, área cubierta por el artículo, etc.* Como puede verse al análisis de contenido así definido comporta una dimensión estadística. Un ejemplo de este tipo de análisis utilizado en investigación documental es el de las *palabras asociadas*, en el cual se buscan los contenidos a partir de las palabras que se repiten en distintos documentos en forma simultánea. No es casual que se lo suela oponer al Análisis de Discurso como antítesis epistemológica, teórica y metodológica: el Análisis de Contenido carece totalmente de postulados relativos a *la naturaleza específica del lenguaje*, tratándolo como un mero vehículo de informaciones. Basándose en la teoría de la comunicación, este enfoque prioriza la posibilidad de realizar observaciones reproducibles y acumulables. Lo que se busca es, como en las disciplinas experimentales, *minimizar la dependencia de los resultados con respecto al punto de vista del investigador*.

Los límites intrínsecos del Análisis de Contenido – por lo menos en su versión clásica – radican, en la manera en que este enfoque *confunde la formalización de las tareas de investigación con una supuesta objetividad científica*. Contrariamente a lo que postula una visión positivista, formalizar es enunciar y argumentar las decisiones metodológicas, decisiones que son siempre locales, parciales y, en última instancia, arbitrarias. Esto es, "objetivar" las reglas analíticas quiere decir construirlas como producto cognitivo siguiendo las "meta-reglas" del lenguaje académico, y no discernirlas como verdades trascendentes.

(3) El *Análisis Estadístico de Datos Textuales* se inscribe de manera general en la tradición francesa del Análisis de Discurso, pero constituye un enfoque muy especializado en el que se procesa lo escrito como un conjunto de *unidades mínimas de sentido* cuyas propiedades pueden ser inferidas mediante *algoritmos matemáticos*.

El Análisis Estadístico de Datos Textuales se distingue de todos los demás enfoques por su rigor operacional: *no se toma ninguna decisión analítica antes de someter el texto a los protocolos lexicométricos*.

Los análisis lexicométricos¹⁷ (estadística textual), se encuentran basados en las técnicas estadísticas desarrolladas por la Escuela Francesa de Análisis de Datos, principalmente por Benzécri en el año de 1973. Los métodos lexicológicos constituyen un conjunto de técnicas estadísticas diseñadas para medir y analizar el vocabulario que conforma un corpus y su particular estructuración. (Cabrera Varela, 1987)

El uso de formalismos y el trabajo con frecuencias son comunes al Análisis Estadístico de Datos Textuales y al Análisis de Contenido – ambos se pretenden métodos de tipo científico –, pero los dos enfoques difieren fundamentalmente en lo que hace a la *concepción de lo textual*.

- El Análisis de Contenido clasifica y contabiliza las unidades de significación en función de una grilla temática "universal", produciendo así un índice de la información transmitida en un mensaje determinado.

- El Análisis Estadístico de Datos Textuales, por el contrario, se focaliza en *las relaciones que se tejen entre las unidades léxicas*, basándose en consideraciones teóricas de la lingüística "distribucional".

(4) El Análisis de Datos Cualitativos es el más reciente de los enfoques que estudian los objetos de lenguaje, cuyo objetivo es *hacer emerger el sentido de lo dicho y lo escrito* sin aplicar categorías exteriores o previas a la observación.

Frente a la obsesión por "medir", por identificar "variables", por descubrir "leyes", ciertos investigadores buscaron desarrollar una perspectiva más acorde con el carácter complejo y sutil de la actividad y el pensamiento humanos. Interesados en la manera en la que los individuos crean y atribuyen significaciones, los investigadores "cualitativistas" observan a los actores en situaciones concretas y recuperan sus propias modalidades de expresión. Recurriendo a estrategias como la "*comprensión*" o la "*observación participante*", intentan capturar los *procesos de interpretación que guían la conducta de los individuos*.

El estudio del lenguaje es, entonces, central para este enfoque, ya que sólo la lectura minuciosa de lo que dicen o escriben los actores permite al investigador reconstruir los diferentes "*universos vivenciales*".

¹⁷ La lexicometría, se refiere a procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y operar algún tipo de análisis estadístico a partir de los resultados de tales recuentos. Se recurre a la cuantificación de los textos desde el primer momento, sin que medien operaciones de codificación previas. Permite facilitar en primera instancia el análisis sintáctico a través del procedimiento de listas de palabras y profundizar en el aspecto semántico a través de la exploración de grandes dimensiones de significado.

Es así que, de alguna manera, el Análisis de Datos Cualitativos comienza a ocupar en el ámbito anglosajón el espacio que corresponde al Análisis de Discurso en el contexto europeo continental y latinoamericano. Con diferencias fundamentales, estos dos enfoques convergen en la voluntad de tratar lo simbólico como un dominio clave de lo social, en el que se construyen y transforman las maneras de pensar y de hacer.

El Análisis de Discurso – con su extrema diversidad de corrientes y especialidades – ocupa, en la universidad francesa y sus áreas de influencia, la mayor parte del terreno textual, mientras que el Análisis Estadístico de Datos Textuales constituye un sector muy acotado y, hasta diría, aislado del resto. Por otra parte, el Análisis de Contenido es comúnmente considerado como una técnica específica, más que como un enfoque general, mientras que el Análisis de Datos Cualitativos se pretende una perspectiva integral.

2.1 DESCRIPCIÓN GENERAL DE LAS TECNOLOGÍAS DE ANÁLISIS DE TEXTOS.

Léxicometría.

Las tecnologías del lenguaje son el conjunto de conocimientos y medios involucrados en el tratamiento del lenguaje humano. (Villaseñor, Aurelio, & Montes, 2003).

Las tecnologías del lenguaje, de habla y texto, procesan o producen expresiones en estos dos tipos de formas del lenguaje. A pesar de esta división, el lenguaje tiene aspectos que son compartidos entre el habla y el texto tal como los diccionarios, la gramática, significado de las oraciones, etc.

Por otro lado una gran parte de las tecnologías del lenguaje no pueden reducirse únicamente a las tecnologías del habla o del texto. Entre esas tecnologías se encuentran las que ligan el lenguaje al conocimiento. No se sabe cómo el lenguaje, el conocimiento y el pensamiento están representados en el cerebro humano; sin embargo, las tecnologías del lenguaje proponen sistemas formales de representación que ligan el lenguaje a conceptos y tareas del mundo real.

El lenguaje empleado por un actor en un momento particular, en un «Aquí y un Ahora», esto es, una actualización particular de aquél en un habla, constituye una acción social realizada en un marco social determinado, en una cultura determinada. Así, pues, cada escrito, además de su significación lexicográfica y gramatical, el momento cultural desde el cual se genera y el momento biográfico de quien la genera. (Cabrera Varela, 1987)

El lenguaje adquiere las connotaciones propias de la subcultura en que se usa y del momento particular en que se emplea. Pero, aún dentro de este marco, cada palabra, cada giro, está enriquecido por la experiencia común de quienes, con su empleo reiterado, lo actualizan constantemente.

Esta experiencia común, restringida a grupos sociales específicos, es la que hace evolucionar al lenguaje de la vida cotidiana a través de sus diversos momentos semánticos. Toda la historia del grupo lingüístico se refleja en su manera de decir las cosas. (Polanco, 2001)

2.2 DESARROLLO DEL ANÁLISIS DE DATOS TEXTUAL

La conjunción entre el análisis textual y el análisis de datos multivariante sufre un continuo desarrollo. A partir de mediados de los ochenta el Análisis de Correspondencias se ha ido haciendo cada vez más popular. A este respecto, la publicación en inglés de las obras de L. Lebart, Morineau y Warwick, (1984) y de M. Greenacre, (1984) ha sido probablemente decisiva.

El Análisis de Correspondencias y en general los métodos de análisis multivariantes se han ido desarrollando conjuntamente en múltiples áreas. A partir de 1990 comenzaron a celebrarse las Journées Internationales d'Analyse Statistique des Données Textuelles (JADT). A la primera celebrada en Barcelona le siguió la segunda celebrada en Montpellier en 1993, en 1995 se celebró la tercera en Roma y continúan de modo regular hasta nuestros días.

En estos años, se produce el desarrollo de métodos y sus aplicaciones a grandes conjuntos de datos procedentes de encuestas socioeconómicas, entrevistas, investigaciones literarias, de textos políticos, archivos históricos, bases de datos documentales, etc. L. Lebart, que años antes había comenzado a investigar en el campo de respuestas libres a cuestiones abiertas en encuestas, desarrolla junto a A. Morineau un primer módulo de tratamiento de textos (1984). Posteriormente M. Bécue Bertaut presenta en la Facultad de Informática de Barcelona su tesis doctoral titulada: Un sistema para el Análisis de Datos Textuales. (Bécue, Análisis de Datos Textuales, Métodos estadísticos y algoritmos, 1991)

La autora investiga de forma sistemática los métodos lexicométricos y estadísticos utilizados en el análisis textual y se centra en un campo privilegiado: el del tratamiento de respuestas abiertas de encuestas y de su relación con respuestas e información cerrada.

Actualmente se progresa en los análisis de respuestas libres a cuestiones abiertas y su relación con el resto de la información recogida en la encuestas. La aplicación de métodos de clasificación es cada vez más frecuente.

Así, la utilización conjunta de análisis factoriales y de clasificación sobre los primeros factores es de gran interés. El método de Clasificación Ascendente Jerárquica comúnmente utilizado (criterio de Ward generalizado) se basa en cálculos de distancia entre elementos de base tomados dos a dos. En general la distancia aplicada es la distancia chi-cuadrado entre perfiles al igual que en Análisis de Correspondencias.

La complementariedad entre las dos técnicas estudiadas se señala que es muy recomendable (Lebart & Salem, 1994) en análisis descriptivos y exploratorios de datos procedentes de grandes encuestas o complejos como los lexicales. La clasificación puede versar sobre los términos o las formas textuales de una tabla lexical efectuando un análisis directo de las respuestas o documentos o puede efectuarse sobre las variables de caracterización de los individuos encuestados o incluso sobre otras cuestiones cerradas en el cuestionario.

En todos los casos los elementos no activos en el análisis pueden proyectarse como suplementarios para ilustrar la descripción de los planos factoriales y clases. Las coordenadas de estas variables o modalidades suplementarias pueden valorarse mediante los “valores-test”, que proporcionan una medida de su significación estadística

2.3 ANÁLISIS ESTADÍSTICO DE DATOS TEXTUALES

Son métodos antiguos, híbridos entre el análisis multidimensional de datos y la Lexicometría¹⁸. El mismo permite la exploración y análisis de textos.

Al igual que la lingüística, el análisis de discursos, el análisis de contenido, se interesan en el análisis de textos; estas “tecnicas” analizan y describen el conocimiento y las relaciones a partir de los documentos disponibles.

El uso del ANÁLISIS ESTADÍSTICO DE DATOS TEXTUALES permite al investigador incorporar la identificación y clasificación de las principales temáticas, a través del ranking de las formas textuales más frecuentes (Feldman, 2002)

Los métodos estadísticos lexicométricos se proponen como sistemáticos, en el sentido de que cuentan la presencia de las palabras sin una selección a priori, estos métodos son exhaustivos, porque trabajan a partir de todo el texto y por lo tanto, son métodos que permiten una mayor objetivación, o por lo menos posibilitan retrasar la introducción de la subjetividad hasta una fase más tardía del

¹⁸ Lexicométricas, de clasificación y análisis factorial de correspondencias. El modelo ha sido desarrollado por Ludovic Lebart y André Salem (Statistique Textuelle, Dunod, Paris, 1994).

trabajo. Los mismos permiten analizar un texto en su generalidad y en su particularidad.

El enfoque cuantitativo que brinda el análisis de datos textuales posee, entre otras, las siguientes ventajas:

- Sus resultados expresados en números brindan un elemento adicional para el análisis de los textos (como es este el caso de las entrevistas)

- Permite realizar comparaciones entre corpus sin necesidad de reelaboraciones, traducciones, etc.

Todo esto sin perder en ningún momento la estructura, la riqueza y la potencialidad del texto originario.

2.4 MARCO DE REFERENCIA

La idea de tratar textos con métodos estadísticos no es nueva. El equipo de Saint-Cloud ha trabajado por años lo que han denominado estadística léxica, produciendo importantes investigaciones en el marco del análisis factorial de correspondencias, con objetivos muy similares: "encarar el problema del discurso en forma sistemática, muy representativa tanto de las posibilidades como de los límites de la utilización de los instrumentos estadísticos" (Maingueneau, 1989)

En los primeros textos se aplicaban los métodos estadísticos elaborados para tratamiento de variables continuas. Sin embargo, los principales y más interesantes resultados se obtuvieron aplicando los métodos de análisis multidimensional de datos para variables nominales.

El desarrollo de las técnicas de la estadística textual ha hecho que el ANÁLISIS ESTADÍSTICO DE DATOS TEXTUALES se haya constituido en una herramienta interdisciplinar, integrada por: la estadística, el análisis del discurso, la lingüística, la informática, el procesamiento de encuestas y la investigación documental; entre otras.

Se utiliza cada vez más en diversos campos de las ciencias sociales: historia, política, economía, educación, sociología, psicología, etc. También, se ha utilizado en el análisis de los discursos sociales, en la investigación del consumidor, del ciudadano, en el análisis de tendencias políticas, etc.

El análisis de datos textuales se complementa con otros métodos lexicométricos y técnicas clásicas como el estudio del vocabulario, las concordancias, las formas características, etc. (Lebart & Salem, 1998) (Bécue, Análisis de Datos Textuales, Métodos estadísticos y algoritmos, 1991) (Etxeberría, García, & Rodríguez, 1995).

2.5 LA DISTRIBUCIÓN DE FRECUENCIAS Y EL CONTENIDO

Éste es el principio fundamental de la lexicometría o estadística lexical: un corpus será representado por la lista de todas las formas lexicales acompañadas, cada una, de un efectivo numérico: su frecuencia de empleo.

La importancia de las medidas de frecuencias de un corpus radica, al menos, en cuatro puntos.

- Primero, la medida de frecuencia es, sobretodo, una herramienta estadística básica para la descripción cuantitativa (Lebart, Salem, & Bécue, 2000) o sea, representa una forma de análisis primordial para los enfoques descriptivos cuantitativos
- Segundo, debido a que la frecuencia como cálculo está matemáticamente relacionada con la estimación de la probabilidad de una unidad en un conjunto de datos, la medida de frecuencia de unidades se encuentra en la base de los modelos estocásticos del lenguaje (Cadenas de Markov y N-gramas)
- Tercero, la frecuencia como noción es muy productiva en cuanto atraviesa los distintos niveles de análisis lingüístico: puede corresponder a un grafema, un morfema, una forma (word form), a una clase gramatical (POS), a un tipo léxico (type), a un caso (token), a una estructura sintáctica, etc.
- Por último, a partir del cálculo de frecuencias se puede observar en distintos niveles el grado de "comunalidad" (commonality) o especificidad (specificity) entre dos o más corpora (Kennedy, 1998)

Suele conocerse, este instrumento, como "index lexical"; será "jerárquico" cuando las formas estén ordenadas por frecuencias decrecientes, o "alfabético". El fragmento de index lexical jerárquico, en el tratamiento exploratorio del discurso, permite establecer una primera imagen sintética del contenido del corpus estudiado. (Pérez Hernández M, 2002)

En razón de que todo discurso emplea, por un lado, un número limitado de formas con frecuencia elevada y, por otro lado, una cantidad importante de formas con escasa frecuencia, se hace posible construir un inventario conciso de palabras que representan una gran parte del repertorio lexical de un corpus.

La frecuencia de palabras se asocia generalmente al nombre de Zipf por haber establecido una ley fundamental (Zipf, 1949) que afirma que, en cualquier texto, el producto de la frecuencia de cada palabra por su rango es constante. Después de Zipf la lingüística moderna ha acumulado una enorme cantidad de material procedente de la observación y descripción de hechos, relaciones, leyes y comparaciones que ha desbordado la capacidad de tratamiento de los problemas por métodos tradicionales y ha elevado a objeto de culto, de forma a veces temeraria, la metodología basada en modelos estadísticos, que, supuestamente,

proporcionan un amplio caudal de interpretaciones a cambio de no demasiada exigencia de conocimientos numéricos.

Shannon desde la teoría matemática de la comunicación, junto a la demanda de nuevos diccionarios más cercanos a la realidad y necesidades de una sociedad cada vez más global, contribuyen en gran manera a la difusión y conocimiento de experiencias, ensayos y formulaciones relacionadas con la estructura y enseñanza de las lenguas y la transmisión de datos. Los nombres de Simón de Laplace, Bayes, Markov o Sharman, entran de lleno en el ámbito lingüístico donde el auge del ordenador y la oferta de potentes programas informáticos, capaces de dar respuesta en segundos a complicados algoritmos matemáticos, colaboran a su desarrollo y favorecen la aparición de nuevas disciplinas.

Despegan la lingüística estadística y la lingüística matemática, que, a su vez, sirven de apoyo para la solución nuevos problemas relacionados, a modo de ejemplo, con la indización y localización en Internet.

Obviamente, no es posible efectuar inferencias sólidas a este nivel de tratamiento pues lo que se percibe es sólo el sistema de "preferencias" lexicales del locutor (sin duda, las palabras "raras" o ausentes contribuyen también al "sentido" del texto); lo que se observa en los rangos superiores del index jerárquico es el conjunto de "puntos de densidad" del discurso, sobre todo en cuanto a su matriz referencial global (pues, dejando de lado las formas funcionales, predominan los sustantivos que vehiculan la "información" intencional). Asimismo, los problemas de polisemia¹⁹ hacen que sea necesario relativizar toda interpretación hasta no verificar en contexto el uso de los términos. (Pérez Hernández M, 2002) Las palabras tienden a mantener en todo contexto un cierto "núcleo" semántico estable (lo que posibilita la existencia de diccionarios). Por ejemplo, cuando un político pronuncia la palabra "patria", se está refiriendo casi con seguridad al colectivo nacional de pertenencia identitaria. Evidentemente, la "connotación" podrá variar de locutor a locutor e incluso de enunciado a enunciado; empero, se tratará en todos los casos de la designación de una "zona" referencial específica²⁰. Sin duda, el trabajo con frecuencias léxicas se vuelve mucho más interesante cuando se establecen contrastes entre varios corpus o entre las diversas partes de un mismo corpus. En el index lexical jerárquico, la frecuencia sólo sirve para dar lugar a un gradiente: el empleo de una palabra es "significativo" cuando su posición jerárquica en la lista indica la existencia de una "preferencia" con respecto a las otras opciones lexicales de las que el locutor disponía (por ejemplo,

¹⁹ Se llama polisemia a la capacidad que tiene una sola palabra para expresar muy distintos significados. Al igual que la homonimia, en el caso de la polisemia se asignan varios significados a un solo significante. Pero, mientras la homonimia se produce por coincidencia de los significantes de diversos signos, la polisemia se debe a la extensión del significado de un solo significante

²⁰ El caso del uso metafórico de las palabras es mucho más complejo y requiere un tratamiento particular.

en un discurso presidencial, Santos "prefirió", en general, el término "país" al término "patria" para designar al colectivo nacional)²¹.

También ciertos aspectos de la enunciación pueden ser indagados a través de un acceso lexical. Por ejemplo, la observación de las frecuencias de empleo de los pronombres personales y posesivos de la primera persona. El estudio de las formas verbales conjugadas en primera persona puede aportar otros elementos relevantes en lo que concierne a la enunciación.

2.6 LAS CONCORDANCIAS Y LA CONTEXTUALIZACIÓN

Tal como afirma Stubbs (Stubbs, 2001), las concordancias no son un invento de la lingüística de corpus en su versión computacional, sino que tienen una larga data en los estudios bíblicos o el estudio del uso del lenguaje en autores clásicos como Shakespeare.

El análisis de un corpus, a partir de las concordancias, permite conocer el contexto en el que se han manifestado formas o lemas que despiertan nuestro interés.

Hasta ahora, se piensa en el texto como una sola secuencia lineal (una frase sigue inmediatamente a la otra y así sucesivamente) en la cual cada unidad léxica ocupa una posición determinada y exclusiva.

Cada unidad tiene un "contexto" propio y único: el de las unidades que se hayan a su derecha y a su izquierda. "Contextualizar" una forma léxica es observar de manera sistemática los contextos sintagmáticos de todas sus ocurrencias.

Una concordancia es, en términos generales, un sintagma²² (que puede coincidir con la frase o con una expansión fija de posiciones a derecha e izquierda) que contiene la ocurrencia de una o varias formas lexicales seleccionadas (y, eventualmente, cumple con uno o varios criterios suplementarios formales o de contenido).

²¹ La "categorización de los datos textuales" – procedimiento que permite, por ejemplo, la creación de unidades de clasificación lexical – es un aspecto clave en este tipo de enfoque. Se trata del conjunto de operaciones que apuntan a superponer a las unidades del discurso (ítems lexicales o supra-lexicales) una o varias grillas de codificación de valor descriptivo o analítico. El objetivo de la categorización es calificar los elementos del texto adscribiéndoles información extra-textual (datos sobre el locutor y las circunstancias de la enunciación) y/o clasificándolos según principios de homogeneidad (morfo-sintáctica, semántica, temática, tópica, etc.).

²² Sintagma, palabra o grupo de palabras relacionadas entre sí y que forman una unidad funcional. Es un tipo de constituyente sintáctico formado por un grupo de palabras que forman otros sub-constituyentes, al menos uno de los cuales es un núcleo sintáctico

El uso de concordancias permite indagar el universo de significaciones que se asocia a una determinada forma léxica. Al extraer del corpus todos los enunciados que contienen cierta palabra (o un conjunto de palabras), el analista puede observar regularidades en el tratamiento que el locutor hace de un tema en particular. También puede efectuarse un análisis de co-ocurrencias: se trata del tratamiento lexicométrico del contexto sintagmático de una palabra tomada como "polo". Se buscará así detectar asociaciones significativas desde el punto de vista estadístico.

Este tipo de procedimientos son esencialmente de naturaleza exploratoria y descriptiva; es evidente que el locutor puede tematizar la palabra sin emplearla o incluso puede tejer una trama de sentido alrededor de ésta sin establecer relaciones sintagmáticas identificables a primera vista.

2.7 TÉCNICAS DESCRIPTIVAS DE ANÁLISIS DE DATOS MULTIVARIANTES

El análisis de datos textuales se apoya en los métodos de análisis estadístico de tablas de datos, concretamente el análisis de correspondencias y el análisis de clasificación.

El análisis de correspondencias de una tabla proporciona una visión gráfica, simplificada, de la información que contiene y pone en evidencia las diferencias entre los distintos perfiles léxicos.

Las representaciones permiten situar a los textos individuales en un espacio determinado por las palabras, permitiendo comparar los perfiles localizando aquellos más semejantes y las palabras causantes de estas semejanzas. Del mismo modo, se detectan los más diferentes explicando sus causas.

La representación de las formas permite estudiar las proximidades entre aquellas formas que son utilizadas simultáneamente por los mismos textos, es decir, estudiar los contextos.

Alternativamente gracias a una utilización conjunta de la información textual y no textual, se verá cuales son las características objetivas de los individuos asociados a un tipo de vocabulario. Se puede ver que si un mismo contenido semántico se expresa con forma distinta según el grupo socioeconómico, sexo, edad, etc.

Con este tipo de análisis, la presencia de palabras herramienta (preposiciones, artículos, conjunciones, pronombres, etc.), está en principio perfectamente justificada si estas palabras caracterizan a algunas categorías, o su posición próxima puede ser interesante para interpretarlas. Si por el contrario, su reparto es aleatorio y se aproximan o se sitúan en la parte central del gráfico, interesa eliminarlas.

Por otra parte, la presencia próxima de varias formas de un mismo verbo, puede constituir una herramienta de validación²³ y justificar la lematización.

El agrupamiento es un proceso de división de un conjunto de datos u objetos en un conjunto de subclases significativas, llamadas clusters.

Las técnicas de agrupamiento de datos es una actividad humana muy importante. Esta actividad usualmente forma las bases del aprendizaje y del conocimiento. Éstas pueden ser usadas eficientemente para encontrar los vecinos más cercanos de un documento, para mejorar la calidad de sistemas de recuperación de información, en la verificación de la homogeneidad de un corpus textual, en el resumen de colección de documentos y en la categorización de términos, entre otros.

Al agrupar los objetos de un conjunto de datos, se requieren algunas medidas para cuantificar el grado de asociación entre ellos. Con este propósito, se pueden utilizar distancias, o medidas de similitud o disimilitud. Algunos algoritmos de agrupamiento tienen un requerimiento teórico para el uso de una medida específica, pero lo más común es que el investigador seleccione qué medida utilizará con determinado método.

Existen varios tipos de técnicas de agrupamiento, entre ellas, técnicas de agrupamiento incompleto o heurístico, técnicas de agrupamiento duro y determinista, técnicas de agrupamiento duro y con solapamiento, técnicas de agrupamiento probabilísticas, técnicas de agrupamiento borroso, técnicas de agrupamiento jerárquico, técnicas de agrupamiento basadas en funciones objetivos y técnicas de estimación de grupos (Höppner, Klawonn, Rudolf, & Runkler, 1999).

En el agrupamiento duro y determinista se asigna cada dato exactamente a un cluster de modo que la partición de clusters defina una partición ordinaria del conjunto de los datos. Mientras que en el agrupamiento duro y con solapamiento cada dato será asignado al menos a un cluster, o puede ser simultáneamente asignado a varios clusters.

Los algoritmos de agrupamiento borroso puro trabajan con grados de pertenencia que indican en qué medida un dato pertenece a los clusters.

La suma de las pertenencias de cada dato a todos los clusters es igual a uno (Höppner, Klawonn, Rudolf, & Runkler, 1999). Algunos ejemplos de algoritmos que

²³ El Análisis Factorial de Correspondencias de la tabla léxica permite detectar cómo algunas palabras, que se podrían considerar a priori semejantes por tener la misma raíz o ser diferentes formas de un mismo verbo, no son empleadas de la misma forma ni en los mismos contextos, por lo que no deberían codificarse en una misma categoría.

aplican técnicas duras y deterministas son: la red de Kohonen (Self-Organizing Maps) para el agrupamiento de documentos (Nürnbergger, Klose, & Kruse, 2001), el algoritmo Autoclass (Larocca, Santos, Kaestner, & Freitas, 2000), el kmeans que es un clásico en el análisis de clusters y también se ha utilizado en dominios textuales.

Todos estos algoritmos tienen ventajas y desventajas que es necesario tener en cuenta al aplicarlos en la solución de un problema. Algunos de ellos requieren que el número de clusters a obtener sea especificado a priori, por tanto es necesario un cierto conocimiento del dominio y en muchos casos, la calidad de la partición final depende de una buena selección de la partición inicial. (Arco, Bello, Meeros, & Pérez, 2006)

Tradicionalmente, para realizar el agrupamiento de los datos, se cuenta básicamente con dos métodos: agrupamiento jerárquico y agrupamiento por partición. En el algoritmo jerárquico, se van creando clusters pequeños, incluso inicialmente con un solo componente, y se van fusionando hasta obtener clusters de tamaño superior; el resultado final es un árbol de clusters conocido como dendrograma, que muestra como los cluster se relacionan unos con otros. Este tipo de algoritmos pueden ser, a su vez de dos tipos: aglomerativos y divisivos. El primero caso corresponde a la estructura de algoritmos jerárquicos que se acaba de describir. Por su parte, en los algoritmos divisivos, se parte de todo el conjunto de datos como un único cluster y, en cada paso, se divide uno de los clusters existentes hasta llegar a un resultado final.

Por otro lado, el agrupamiento por partición es aquel que distribuye los objetos del universo de estudio en grupos (cluster), buscando maximizar alguna medida de similitud entre pares de patrones, entre un patrón y un grupo, y finalmente, entre pares de grupos, de forma que los objetos en un mismo grupo sean muy similares entre sí (cohesión interna del grupo) y los de los objetos de cluster diferentes sean distintos (aislamiento externo del grupo).

3. PROTOCOLO PARA EL ESTUDIO DE TEXTOS

3.1 MÉTODOS DE OBSERVACIÓN, EXPLORACIÓN Y COMUNICACIÓN APLICADOS AL ESTUDIO DE TEXTOS

Una Metodología (reglas de procedimiento) aspira a ligar, en forma válida, una secuencia de actuaciones y alternativas que atañen a la selección del corpus, su segmentación en unidades de análisis y su examen orientado a hacer aparecer la estructuración intrínseca de los conocimientos relevantes.

La metodología propuesta en esta tesis comporta en su realización cuatro fases: (1) adquisición terminológica, (2) control del vocabulario, (3) clasificación de los términos y documentos (constitución de los clusters), (4) descripción conceptual de los clusters y comentario.

El enfoque estadístico (fase 3) proporciona el soporte objetivo para que quien hace el estudio del texto precise cuál es el significado conceptual de los clusters (fase 4). En efecto, la parte más importante del trabajo está consagrada en la práctica al análisis de los clusters con el objetivo de determinar su significación conceptual, y caracterizar en función de esta significación la posición relativa de los clusters.

Las técnicas de análisis de datos que se han tomado de base en esta propuesta, pueden dividirse en el análisis factorial y clasificación. El análisis factorial busca proyectar los datos sobre un espacio de dimensión reducida que guarde buena parte de la información original. Los métodos de clasificación intentan reagrupar los individuos en clases homogéneas respecto a las variables observadas. Se trata de métodos complementarios.

Los métodos de análisis estadísticos multidimensionales complementan las antiguas técnicas de lexicometría²⁴, realizando el tratamiento de los textos considerando una nueva variable léxica, cuyas distintas modalidades serán las formas léxicas o más exactamente, las formas gráficas²⁵ del corpus tratado.

Para la estrategia metodológica, todos los textos a analizar pueden ser codificados con variables cualitativas categóricas o cuantitativas continuas (datos complementarios que describen el objeto observado) y pueden incluir un

²⁴ Bajo el nombre de métodos Lexicométricos se engloba los propios del tratamiento de datos textuales como la creación y edición de glosarios del vocabulario, índices y concordancias, selección de elementos característicos de las distintas partes del documento

²⁵ Se llama forma gráfica a una sucesión de caracteres (en general letras) separados por delimitadores (blancos y signos de puntuación).

identificativo que corresponde a individuo o "casos de estudio" (ej. Respuestas individuales a preguntas abiertas o fragmentos de textos de un libro...). Cada "individuo u observación"²⁶ es un fragmento de texto, el conjunto de todos estos textos forman el llamado "corpus".

El análisis de datos textuales se complementa con otros métodos lexicométricos y técnicas clásicas como el estudio del vocabulario, las concordancias, las formas características, las frases características de las partes del corpus, etc. (Lebart & Salem, 1998) (Bécue, Análisis de Datos Textuales, Métodos estadísticos y algoritmos, 1991) (Etxeberría, García, & Rodríguez, 1995).

En esta tesis en la propuesta para el análisis de textos, qué métodos aplicar, cuándo y cómo (estrategia), depende del tipo de estudio: "estudio de diferencias de textos".

3.2 ESTRATEGIAS METODOLÓGICAS PROPUESTAS

Se ofrece, con esta tesis, un instrumento para abordar la comparación de distintos textos, que opera a partir del recuento exhaustivo de las palabras y los segmentos, pero teniendo en cuenta que los mismos operan dentro de un universo²⁷ que representa el documento analizado.

La propuesta involucra la aplicación métodos, en especial el análisis de correspondencias y la clasificación, a tablas específicas creadas a partir de los datos textuales. Estos métodos se completan con métodos propios del dominio textual como los glosarios de palabras, las concordancias y la selección del vocabulario más específico de cada texto, para así proveer una herramienta comparativa de los mismos.

El principio fundamental es el análisis a través de la comparación. Se busca comparar entre sí el "discurso" de los individuos, o de grupos de individuos con características comunes. En un ámbito literario permite asimismo comparar textos. La comparación implica llegar eventualmente a clasificar a los individuos o a los textos en clases homogéneas en cuanto al vocabulario empleado. También puede interesar clasificar palabras. Estos métodos pueden resumir los textos mediante

²⁶ En estadística se le llaman "individuo estadístico" a la mínima unidad de observación, sobre cada uno de los cuales se ha medido u observado una o varias características (el conjunto de valores posibles para todo los individuos forman una variable).

²⁷ La transparencia del contenido respecto al análisis permite descifrar la información sobre el significado de las palabras, según el objeto de estudio, contenida en el propio corpus. En el momento de interpretar los resultados se reintroduce lo que se conoce de la lengua, de las palabras y de los propios textos. Ello permite transformar lo cuantificable en significativo y volver al contenido al final del análisis.

las palabras y frases más características, concepto que se desarrolla más adelante.

3.3 TIPO DE ESTUDIO

Con este protocolo que se presenta en esta tesis, se pretende trabajar con datos reales y lo más exhaustivos posibles que permitan reproducir con la máxima fidelidad las características del objeto de estudio. Esto implica que hay que recopilar muestras de los elementos que constituyen la realidad que se quiere observar. En su aplicación, los textos a analizar pueden ser: un único texto (ej. una entrevista, un libro, etc.); un conjunto de textos (ej. más entrevistas, artículos de periódicos, etc.).

En definitiva, se propone, en función del objetivo del estudio diferenciar²⁸ el tipo de enfoque, para la ejecución del protocolo de análisis formulado en esta tesis:

3.4 ESTUDIO COMPARATIVO DE TEXTOS

El objetivo es el análisis comparativo de fragmentos de textos.

Estos, pueden provenir de una observación y producción de información textual para estudios que van más allá de las encuestas de opinión, ser constituidos por un conjunto de artículos de periódicos o incluso, encuestas o test psicológicos, textos literarios o científicos.

Frecuentemente, la palabra puede dar información sobre el contenido, indicando qué parte del documento, notable por sus características formales, puede presentar un contenido relevante (Sperber, 1986).

3.5 FASES Y ACTIVIDADES PROPUESTAS

La guía metodológica presentada en esta tesis, permite el análisis léxico de los textos y especialmente la construcción de estructuras de clasificación y categorización que se codifican en forma de tesauros²⁹.

²⁸ El Análisis estadístico de textos tiene aplicaciones en múltiples ámbitos: desde el análisis de respuestas abiertas en encuestas hasta el estudio de corpus de textos literarios, políticos o científicos, sin olvidar la extracción de información de archivos históricos y de bases documentales

²⁹ No en vano tesoro proviene del griego thesaurós: tesoro. La importancia del tesoro consiste en que cada uno de sus términos, se utiliza para denotar un concepto, la unidad semántica básica, que permite expresar una idea.

Secuencia lógica del proceso de análisis de la información en textos que se ajusta a la propuesta formulada en esta tesis.

CORPUS. Creación del corpus a partir de las bases de datos textuales.

BIBLIOMETRIA. Definir y producir los análisis estadísticos que corresponden a los indicadores bibliométricos tradicionales. Se llama "bibliometria", a la aplicación de técnicas estadísticas sobre datos textuales. Generar una información cuantitativa sobre la frecuencia y distribución de los datos, procurando así al analista una información estadística descriptiva acerca de los datos del corpus.

Produce la materia estadística de base, análisis estadístico que se completa con el paso al análisis multidimensional, de la Infometría, mediante la aplicación de técnicas de clasificación conocidas en el análisis de datos.

INDEXACION. Revisar la indización que toma en cuenta la variación en los textos de los términos de un vocabulario. Visualización y la validación, de los resultados.

Esta fase podemos considerarla como aquella del "análisis terminológico" de la lengua en cuestión. La Indización permite analizar la terminología caracterizando el contenido conceptual de los documentos.

Las palabras claves juegan el rol de representar lingüísticamente conceptos. Su objetivo es representar y describir el contenido de los textos, mediante conceptos principales contenidos en ellos (palabras claves) o vocabularios controlados (descriptores, segmentos). Se trata simplemente de una lista de términos (o palabras claves) sin otro orden que la frecuencia de aparición en la colección de documentos.

Es la fase previa y necesaria para la clasificación que supone la existencia de una matriz [datos x palabras claves o términos]. En efecto, contar con datos indexados es una condición necesaria para su clasificación (luego en el módulo INFOMETRIA).

INFOMETRIA³⁰. Ejecución de métodos de clasificación, al mismo tiempo la visualización de los resultados como listas.

³⁰ El término "infometría" se emplea para significar de una manera general los análisis métricos de la información (es decir estadísticos, probabilistas y de análisis de datos), aplicados a la producción de una "información de la información" del tipo "quién hace qué, en cooperación con quién, en donde y cuando". (Polanco, Transformación de la información en conocimiento y del conocimiento en decisiones estratégicas, 2002), el término "infometría" se emplea para significar de una manera general los análisis métricos de la información (es decir estadísticos, probabilistas)

La transparencia del contenido respecto al análisis permite descifrar la información sobre el significado de las palabras, según los textos individuales observados, contenida en el propio corpus. En el momento de interpretar los resultados se reintroduce lo que se conoce de la lengua, de las palabras y de los propios textos. Ello permite transformar lo cuantificable en significativo y volver al contenido al final del análisis.

Análisis multivariado de datos textuales. Aplicación del Análisis Factorial de Correspondencias, sobre las tablas lexicográficas o la Clasificación (Clasificación jerárquica ascendente) de las formas lexicales y textos.

La clasificación que es propuesta en la INFOMETRIA se realiza sin ninguna información previa acerca de las clases a obtener, no se trata de clasificar en función de una taxonomía preexistente en donde el problema consiste en afectar los datos a las clases previamente definidas. Esta clasificación constituye en el análisis de datos un método exploratorio, es decir, que busca descubrir en los datos mismos una estructura de ellos en clases. En principio, las clases agrupan los datos en función de su proximidad o similitud.

Identificación frases modales combinando los resultados del análisis textual con variables sociodemográficas y actitudinales, se obtiene una Tipología de individuos o grupos a partir de respuestas y de textos.

Este trabajo de análisis de la información debe ceder la plaza al experto que dado sus conocimientos y experiencia, asistido por estos indicadores, es capaz de realizar la interpretación de las clases y de sus posiciones relativas en el espacio de conocimiento. La interpretación del experto podrá validar o no, y agregar comentarios pertinentes.

Para que los corpus faciliten la extracción de datos homogéneos y cuantificables de manera que permitan elaborar teorías empíricas, es necesario restringir las diferentes ocurrencias léxicas a ocurrencias formales comunes (unidades estandarizadas); para ello es necesario reducir las variantes a invariantes.

Para la aplicación de métodos estadísticos es necesario cuantificar las características del corpus a estudiar, lo cuál conduce a definir múltiples índices. Entre los distintos índices, se pueden mencionar, por ejemplo, la frecuencia de una determinada palabra, considerada particularmente importante, o el número de veces que un verbo viene seguido de un sustantivo. Se suele utilizar también índices calculados a partir de las longitudes de las palabras, las frases, etc. y de análisis de datos), aplicados a la producción de una "información de la información" del tipo "quién hace qué, en cooperación con quién, en donde y cuando". (Polanco, Transformación de la información en conocimiento y del conocimiento en decisiones estratégicas, 2002)

Sobre los procedimientos de análisis de textos que se plantean en esta tesis, se realizan las observaciones que se explican a continuación.

En el estudio de las tablas de contingencia lexicales, si el analista y experto en el dominio de análisis, lo consideran necesario, es necesario volver a la segmentación para eliminar o seleccionar nuevas formas. Este refinamiento implica adaptar el vocabulario y tablas de acuerdo a la información que se va obteniendo para llegar a los objetivos de estudio. Después de la desambiguación y lematización, e incluso después de seleccionar las formas claves, siempre es necesario realizar un nuevo estudio de la estructura léxica del vocabulario. Estos procedimientos son paralelos y complementarios para obtener una adecuada segmentación del corpus.

Con la Estadística Multivariante se busca principalmente investigar la existencia grupos de unidades de observación que se caracterizan por la utilización de un lenguaje similar, es decir grupos de textos que presentan perfiles léxicos similares. Esto implica que el conocimiento que se obtiene de los individuos está basado en la identificación de si dicen aproximadamente lo mismo, y no en lo que dicen. La lectura de las diferentes salidas del análisis le permite al investigador tener una idea bastante general del conocimiento presente en los textos.

El Análisis de Correspondencias es aplicado, en el marco de la Estadística Textual, a tablas de contingencia en las que se dispone de información sobre la frecuencia en que aparecen determinadas unidades textuales en distintos textos o partes de un corpus textual considerado. Si las tablas de contingencia son construidas agrupando en cada columna a los individuos (tabla léxica agregada) para los cuales una determinada variable adopta el mismo valor (edad, nivel social...), se puede identificar la posición relativa de unos grupos respecto a otros de acuerdo con el vocabulario empleado, y caracterizar la posición de los mismos por las formas próximas en el espacio.

Los métodos de clasificación conducirán a la construcción de clases de individuos homogéneas en cuanto al vocabulario empleado, o a la determinación de grupos de palabras que suelen ser empleadas por los mismos individuos y que delimitan, por tanto, campos semánticos o temáticas conectadas entre sí. La clasificación de los individuos o unidades textuales puede realizarse a partir de las coordenadas de estos tras llevar a cabo un Análisis de Correspondencias. Ambos métodos de análisis pueden ser utilizados conjuntamente, permitiendo una mejor descripción de los datos.

Los valores-test (véase (Lebart & Salem, 1994, pág. 95), (Etxeberría, García, & Rodríguez, 1995)) conforman una herramienta de caracterización³¹ de un eje

³¹ Los valores-test son mayores cuanto mayor sea la coordenada de una categoría suplementaria sobre el factor y cuanto mayor sea el número de individuos que escogen esa modalidad. Bajo la hipótesis nula de que

factorial de un Análisis de Correspondencias a partir de las modalidades de una variable suplementaria o de un grupo o cluster obtenido a través de un método de clasificación.

Para caracterizar un grupo de una clasificación, el valor-test, bajo la hipótesis nula de que la proporción de individuos que escogen una modalidad dentro de una clase es la misma que la proporción de individuos que la escogen en el total de la muestra, sigue una distribución hipergeométrica.

El estudio de concordancia es una herramienta muy importante que se utilizará tanto en la preparación del vocabulario, como en la interpretación de los resultados, para clarificar dudas respecto al contexto de ciertos lemas o palabras claves. En cada etapa o procedimiento se obtienen una serie de resultados con la interpretación y validación estadística (si corresponde) de los indicadores de conocimiento que se van obteniendo, de esta manera al finalizar se podrá integrar en un informe global todas las conclusiones e inferir el “conocimiento” necesario para llegar al objetivo de estudio de los datos textuales.

El objetivo no es la realización de operaciones estadísticas sobre documentos o datos textuales contenidos en bases de datos, sino que los datos van a constituir la materia prima de la cual hay que extraer el conocimiento subyacente, así como la generación de nuevo conocimiento, utilizando en este caso indicadores relacionales y específicos de conocimiento. Los conocimientos se encuentran materializados en textos. Los indicadores son capaces de expresar los conocimientos contenidos en esa cantidad de información textual.

3.6 PROPUESTA PARA EL ESTUDIO COMPARATIVO DE TEXTOS

Tras recopilar, digitalizar y revisar un corpus que contenga información pertinente, completa y actual, se procede a la extracción de términos.

El análisis y agrupamiento es de “textos” o fragmentos de textos. En este tipo de estudio los glosarios son extensos y complicados, para analizar la repartición del vocabulario se sugieren métodos alternativos como el estudio de la asociación de palabras y determinación del vocabulario “típico” y “específico” de cada texto o de un conjunto de textos. Para estos procedimientos es recomendable aplicar algoritmos de lematización y desambiguación para reducir y normalizar el glosario de formas.

la categoría no esté representada mejor que otra u otras sobre el eje factorial, el valor-test sigue una distribución normal tipificada.

- Formar una tabla de contingencia (textos*formas) o sea una “tabla léxica básica” y una tabla de contingencia (Formas*grupos de textos) o sea una “tabla léxica agregada”.
- Aplicación del Análisis Factorial de Correspondencias³², sobre las tablas lexicográficas o la Clasificación Automática (Clasificación jerárquica ascendente) de las formas lexicales y textos.
- Selección de formas características de cada texto, escogidas según un cierto criterio como representantes del texto.
- Obtención de Tipologías o grupos a partir de de textos. Asociación de variables estructuradas, al análisis de las tablas léxicas permitiendo la clasificación según los léxicos empleados y las modalidades escogidas en las variables.
- Representación de la distribución del corpus lexicográfico mediante Mapas preceptuales.

Utilización del Análisis de Correspondencias para la representación gráfica de la información contenida en las Tablas léxicas.

Estudio de la tabla léxica base. Análisis de la tabla léxica (textos individuales * vocabulario). Agrupamiento de textos en función de su vocabulario. Estudio de la tabla léxica agregada. Análisis de la tabla léxica (vocabulario * textos individuales). Agrupamiento del vocabulario según su distribución por lo textos. Análisis de la tabla agregada (vocabulario * grupos de textos). Agrupamiento del vocabulario según su distribución en grupos de textos con base en datos complementarios (por ejemplo fecha, autor, etc).

3.7 DESCRIPCIÓN DEL PROCEDIMIENTO PROPUESTO

A continuación se enseñan los procedimientos apuntados para el uso del protocolo de análisis de textos propuesto en esta tesis.

A. Preparación del corpus

Se indica para aquellos textos (narraciones, artículos periodísticos, informes, etc.) que hayan sido recopilados en soporte papel, deben ser digitalizados. Una vez hecho esto, hay que revisar la ortografía de cada uno de los documentos y homogeneizar su formato.

³² La aplicación del Análisis Factorial en el campo de análisis de datos textuales, se centra, principalmente, en el Análisis Factorial de Correspondencias, algoritmo estadístico desarrollado por Jean Pau Benzécri (Benzécri, 1973). Se trata de un método descriptivo (no explicativo) que se clasifica entre los métodos multivariados de interdependencia y permite visualizar los datos (que pueden ser cualitativos o cuantitativos) mediante la representación de una nube de puntos en un espacio de dimensiones reducidas, en función de las distancias euclidianas entre los puntos

Al ser los textos, datos no estructurados, se hace necesario un preprocesamiento de los mismos, lo que permite la transformación, del corpus. Para la transformación se apunta las operaciones siguientes: la sustitución de las contracciones por sus expansiones, de las abreviaturas por sus formas completas y la eliminación de números y símbolos, las mayúsculas de inicio de oración se pasan a minúscula y verificación de la homogeneidad ortográfica.

Cuando se quieren obtener las formas por orden alfabético, los signos diacríticos³³ se propone indicarlos con un carácter especial después de la letra afectada, ya que, en el caso de las letras acentuadas o con signo diacrítico, el orden lexicográfico incluido por la codificación binaria de los caracteres no corresponde con el orden alfabético.

En el estudio comparativo de textos se pretende reducir el tamaño de los documentos eliminando las partes de los textos que dan poca información sobre su contenido, es decir, que carecen de significado temático.

A.2. Partición del corpus

Se sintetizan las principales actividades propuestas en esta tesis, para obtener los “*individuos estadísticos observados*”, es decir identificar las partes o textos “individuales” que conforman el corpus.

Siendo que el corpus es una sucesión de ocurrencias de palabras y de delimitadores, esta sucesión puede ser particionada de diferentes maneras. Básicamente se habla de dos particiones jerarquizadas a saber: el corpus está compuesto de “oraciones individuales” que se pueden agrupar en “textos”.

En esta propuesta se dice que el documento se descompone en partes que se pueden llamar “textos”. Esta partición puede corresponder a una realidad “a priori”, caso de respuestas abiertas (opinión escrita de alguien), o ser decidida de forma “arbitraria”, caso de fragmentos, frases o párrafos de, por ejemplo, un texto literario o científico (trozos de texto caracterizados o tipificados por datos como el tiempo o autor).

La partición del texto puede ser el resultado de un agrupamiento de partes individuales según un criterio de clasificación, es decir el corpus completo se divide en textos, según las características de los “individuos estadísticos observados” (se reagrupan según variables que tipifican o segmentan los textos,

³³ Un signo diacrítico es un signo gráfico que confiere a los signos escritos (no necesariamente letras) un valor especial. Son diacríticos, por ejemplo: los acentos ortográficos, la diéresis, los signos empleados en el alfabeto fonético, como la oclusión (^) o la nasalización (~), la tilde de la ñ, la cedilla (,), la colita (¸), el carón (ˇ), el breve (˘), el macrón (¯), el anillo (°), y el garfio (‘).

variables que actúan como predictores, etc.) para obtener partes del corpus o subcorpus.

En definitiva, la partición del corpus en textos, puede venir dada a priori, o puede ser el resultado de un agrupamiento según un criterio externo.

El resultado final propuesto será un fichero con la variable léxica: el corpus adecuadamente identificadas sus partes. Los datos textuales se graban con una mínima precodificación: se indica el principio del texto y el principio de cada individuo.

A.3. Preparar base de datos

Se sintetizan las principales actividades propuestas en la tesis, para obtener el Fichero de datos con la variable textual y variables numéricas y cualitativas.

En este procedimiento se plantea crear el fichero base o tabla con la variable léxica. En esta tabla de datos se incluye una o varias variables léxicas (siempre se colocan al final de la base de datos, últimas columnas) y otras variables numéricas y/o cualitativas complementarias.

Nota: Cuando se manejan corpus de gran tamaño suele manipularse, en un principio, dos ficheros separados uno para la variable léxica y otro para los datos complementarios, dependiendo de la herramienta a utilizar, ambos ficheros se fusionan. Para esto, es importante, que ambas tablas tengan el mismo identificador de cada "individuo estadístico" (cada respuesta o parte del texto).

El resultado final propuesto será un fichero con tabla de datos numéricos o cualitativos. Los datos cuantitativos y cualitativos, son valores observados sobre un grupo de individuos o unidades de observación que coinciden con la mínima unidad de análisis del archivo de textos (texto o respuesta individual). Al final del mismo se agrega la variable léxica.

B. Segmentación

El tratamiento preliminar del corpus, propuesto en este procedimiento consiste en identificar las palabras y los segmentos repetidos, contarlos y enumerarlos, en general por orden alfabético y de frecuencia. Se obtiene así una primera información sobre el corpus, además de su longitud y el tamaño de su vocabulario (glosario de formas gráficas).

Siguiendo diferentes criterios, es posible diferenciar diversas unidades que han sido tomadas en consideración por los investigadores: formas gráficas y segmentos repetidos.

En esta tesis se habla de que la unidad de estadística (lo que se va a contar), que servirá de base a las comparaciones debe ser invariante y conservar su identidad en sus distintas repeticiones. Por esto se busca una definición formal que permita atribuir las ocurrencias a una unidad dada.

Para poder efectuar los tratamientos estadísticos sobre el corpus de forma eficiente, se plantea su codificación, es decir, representar cada palabra mediante un entero.

La correspondencia entre una palabra y su traducción numérica se almacena en un diccionario de formas.

B.1. Elaboración de documentos lexicométricos

Se sintetizan las principales actividades propuestas en esta tesis para obtener las unidades de análisis elementales y la Creación de un índice de los textos.

-Obtener las “unidades de análisis elementales”³⁴

Se plantea obtener dos tipos de resultados básicos, uno correspondiente a las formas (palabras) y otro a los segmentos. La palabra será la unidad de base. La segunda unidad estadística considerada es el segmento³⁵ de frase repetido. Esta nueva unidad estadística, permite tomar en cuenta el contexto de las formas. Las formas gráficas se sustituyen por su número alfabético en el vocabulario del corpus.

Formas gráficas

La primera tarea que se apunta, es la determinación de la unidad mínima³⁶ de los cálculos estadísticos, unidad mínima en el sentido de que la descomposición del corpus se hará en estas unidades. La unidad adoptada es la forma gráfica definida como la sucesión de caracteres comprendidos entre dos delimitadores (por ejemplo espacios).

Por lo tanto la forma que surge en forma inmediata es la palabra, sin embargo es necesario remitirse a las palabras tal como vienen escritas, de manera que singular y plural de un mismo sustantivo son dos formas distintas así como las distintas inflexiones de un verbo.

³⁴ Recuento exhaustivo de las palabras y los segmentos en el corpus

³⁵ Compuestos por sucesiones de palabras principales, cuyo tamaño es predefinido (de 8 a 20 palabras) o por una puntuación clara (punto y parte, por ejemplo)

³⁶ Elección de la unidad estadística con la cual se va a trabajar

Las formas gráficas son fácilmente identificables y permiten una segmentación del texto en unidades inequívocamente delimitadas, a diferencia de lo que ocurre cuando el analista procede a segmentar el texto utilizado como criterio de división, por ejemplo, la unidad temática. (Etxeberría, García, & Rodríguez, 1995)

Los segmentos repetidos

Los Segmentos Frecuentes: es una secuencia de N palabras consecutivas. Son los que aparecen en el texto un mínimo de frecuencia o más veces. Son unidades más largas, compuestas por varias formas o lemas, dotadas de un sentido propio y que no siempre se deduce de las palabras que entran en su composición, si éstas se consideran aisladamente.

Son interesantes porque permiten extraer, mediante un algoritmo recursivo, las combinaciones de palabras más frecuentes (después de extraer las palabras vacías y de aplicar lematización obviamente). Es una manera sencilla de saber por ejemplo de qué está hablando un determinado texto; Solo viendo los mayores segmentos repetidos podemos tener una buena idea de lo que habla.

Esta unidad de análisis fue introducida por Salem (1987). Todo el tratamiento estadístico aplicable a las palabras, puede generalizarse a los segmentos. El análisis de los segmentos, permite tomar en consideración el contexto de las palabras. Es decir, permiten considerar el entorno sintagmático en el que las palabras son empleadas y pone de manifiesto secuencias que se repiten y que en la lectura inicial del texto posiblemente no hubieran sido detectadas.

En este procedimiento, para la codificación y segmentación del corpus, se plantea definir dos tipos de delimitadores, parámetros que necesita el algoritmo para obtener formas y segmentos repetidos.

Los delimitadores más comunes son el “espacio”, la “coma” y el “punto”. Son considerados como delimitadores débiles, los dos primeros y el tercero como delimitador fuerte. Se puede elegir otros delimitadores, es una decisión que depende del tratamiento deseado.

-Creación de un índice de los textos (partes del corpus)

Una vez segmentado el texto se está en condiciones de construir diversos documentos lexicométricos, consiste en una reorganización de las unidades contenidas en el texto siguiendo determinados criterios. El más simple es el glosario o vocabulario del texto, que puede ser presentado en orden e frecuencia decreciente o en orden alfabético cuando la frecuencia es igual.

Se plantea asociar a cada forma a las coordenadas de sus ocurrencias en el corpus, se obtiene el índice del corpus. El índice puede ser en orden lexicográfico

(índice lexicográfico) o en orden de frecuencia (índice jerárquico). El índice permite localizar cada una de las ocurrencias en el corpus. Por ejemplo, puede ser interesante listar todos los contextos de una misma palabra.

El “índice de los textos” indica los términos usados en cada corpus y sus frecuencias de ocurrencia. Su representación es mediante un archivo invertido, es decir, una estructura de datos que consta de un diccionario y una lista invertida instrumentados a través de tablas de hash³⁷.

Estas tablas almacenan la información en posiciones pseudo-aleatorias, así que el acceso ordenado a su contenido es bastante lento. Otras estructuras como árboles binarios³⁸ auto-balanceables, son más lentos en promedio pero la información está ordenada en todo momento.

En el diccionario se almacenan todos los términos extraídos, junto con su frecuencia total de ocurrencia. En la lista invertida se almacena, para cada término, una lista dinámica de las colecciones (podrían ser más de dos) en los que el término fue encontrado y la frecuencia de ocurrencia en cada una de ellas.

A partir del índice construido, una frecuencia es asignada a cada uno de los términos. Esta frecuencia indica el número de ocurrencias del término k en cada texto (en caso de comparar dos textos, $i = \{1,2\}$). Con base en estas frecuencias se construye una distribución de probabilidad de los términos en el corpus i , donde:

$$p_k^i = f_k^i / \sum_{j=1}^n f_j^i$$

Expresa la probabilidad de ocurrencia del término k en el corpus i , y n indica el número de términos existentes en el índice.

El perfil léxico es un vector cuyos componentes son las frecuencias de cada una de las formas utilizadas por un individuo (texto individual) o un grupo de individuos. Una vez conocido el vocabulario del texto, interesa conocer los perfiles de las frecuencias de las formas y especialmente las diferencias entre ellos.

³⁷ Una tabla hash es una estructura de datos que asocia llaves o claves con valores. La operación principal que soporta de manera eficiente es la búsqueda: permite el acceso a los elementos (palabras) almacenados a partir de una clave generada. Funciona transformando la clave con una función hash en un hash, un número que la tabla hash utiliza para localizar el valor deseado

³⁸ Un árbol binario de búsqueda auto-balanceable o equilibrado es un árbol binario de búsqueda que intenta mantener su altura, o el número de niveles de nodos bajo la raíz, tan pequeños como sea posible en todo momento, automáticamente

El resultado final propuesto es un listado alfabético de formas y segmentos, acompañadas de su frecuencia absoluta, así como de su longitud; al final del mismo se aportan también los siguientes datos estadísticos: número total de formas, número total de formas distintas, porcentaje de formas distintas..

B.2. Análisis de la estructura léxica del vocabulario

Se sintetizan las principales actividades propuestas en la tesis para obtener: Diversidad, originalidad, especialización y crecimiento del vocabulario.

El glosario pone en evidencia la frecuencia de repetición de las palabras. Un complemento propuesto para la lectura de las frecuencias, es el estudio de la “repartición de las palabras”.

Por ejemplo una palabra empleada 50 veces en un corpus de 50000 palabras, puede aparecer una vez cada mil palabras, con la regularidad de un reloj; en este caso, la repartición es estable. Al contrario si las 50 repeticiones pueden aglutinarse en un punto del corpus, entonces se trata de una palabra localizada.

-Creación de un Índice para medir la repartición de palabras.

Se plantea realizar una medida y comparación de la riqueza del vocabulario del documento. Las palabras se suelen diferenciar en “semánticamente llenas” y “gramaticales o funcionales”. Esta diferenciación conlleva a decir qué palabras tienen una utilización específica diferenciada en los textos del corpus, o qué palabras contienen algún tipo de información, o qué palabras ayudarán, en el momento de interpretar los resultados, a entender el funcionamiento del texto estudiado. A priori es difícil determinar, para un corpus, cuales palabras son de una categoría u otra.

Para una palabra de frecuencia igual a F en un corpus de longitud N , dicho índice, que se calcula a partir de la longitud de los intervalos que separan las reparticiones y varía de 0 a 1. Se puede considerar como una aproximación a la probabilidad de encontrar esta palabra en cualquier parte del corpus de longitud igual a N/F (frecuencia relativa). Un valor próximo a 1 indica que la palabra se emplea de manera habitual, a lo largo del corpus, un valor próximo a 0, es la marca de una utilización circunstancial, localizada de la palabra. (Bécue, Análisis Estadístico de Textos, 1997)

La estabilidad de las palabras depende de la naturaleza del corpus, un factor que interviene es el uso, impuesto por la lengua, de ciertas palabras herramientas (gramaticales o funcionales), de hecho gran número de palabras estables son de este tipo, pero no son ni las únicas, ni las regulares.

Para inferir información a partir de este tipo de listas, se sugiere retornar al texto, para esto, el estudio de las concordancias, es decir listado sistemático de todos los contextos de una palabra es de gran ayuda (estudio que se realiza en la siguiente fase para depurar el glosario).

-Estudio de la Gama de frecuencia

George Kinsley Zipf³⁹ supone que el ser humano trata siempre de minimizar el esfuerzo necesario para lograr sus objetivos, lo que, en circunstancias donde le es permitido escoger alternativas, le lleva a inclinarse por aquellos procesos que resultan en el menor consumo de energía. Esto significa la existencia del principio del menor esfuerzo (Zipf, 1949) en la conducta humana y hace posible que, casi siempre, sus patrones de actuación puedan ser analizados de acuerdo con este principio.

De igual modo que cuando busca la unión de dos puntos tiende hacia la línea recta, cuando escribe, cuando habla, tiende hacia la mayor economía de palabras. Basándose en estas observaciones Zipf formula su ley que relaciona frecuencia y rango.

La gama de frecuencia ofrece una imagen de la estructura léxica del corpus. En general, se propone intentar determinar las regularidades existentes en la gama de frecuencia para obtener leyes o modelos que sirvan de referencia. Las desviaciones respecto a las regularidades podrán en evidencia los rasgos propios de la estructura de un corpus particular. (Bécue, Análisis Estadístico de Textos, 1997).

El corpus se caracteriza por la frecuencia máxima o frecuencia de la palabra más frecuente, así como por la distribución de las formas según su frecuencia representada por el histograma de los efectivos por frecuencias.

A partir de esta información se puede construir la gama de frecuencias, que es la tabla que asigna a cada frecuencia el número de palabras que la asumen, desde 1 hasta la frecuencia máxima.

En definitiva el procedimiento propuesto para cuantificar ambos conceptos (frecuencia y rango) empieza por contar las veces que se repite cada palabra en un texto. Obtiene el indicador frecuencia dividiendo el resultado de cada recuento entre el total de palabras del texto. Se ordenan ahora todos los términos siguiendo el orden decreciente de las frecuencias resultantes.

Se denomina rango de la palabra al lugar que ocupa cada término en ese listado. Si en un texto la palabra de más frecuencia es "el", en la lista ocupará el primer

³⁹ (1902-1950), profesor de Filología en la Universidad de Harvard

lugar y, por tanto, tendrá rango uno. Si el artículo "la" tiene el segundo valor de frecuencia tendrá rango dos.

Cuanto mayor es el rango menor es la frecuencia con la que aparece en el texto. Un rango alto se sitúa en la parte baja de la lista y eso significa menor frecuencia. Esta dependencia actúa en forma inversa porque disminuye a medida que el rango aumenta. Si f denota frecuencia y r rango, f depende de r como $1/r$.

Este resultado se llama ley de Zipf de rango-frecuencia (Zipf, 1949) $r \times f = C$ donde, r es el orden de la palabra en la lista (rango), f es la frecuencia y C es la constante para el texto.

La ley de Zipf vincula la frecuencia de ocurrencia de una palabra y el número total de palabras conocidas o utilizadas; esto es, la amplitud de vocabulario. Si ordenamos las palabras de un texto de la más común a la más rara, la frecuencia decrece exponencialmente, esta es una característica del lenguaje humano.

La ley de Zipf expresada formalmente dice que la frecuencia de una palabra en un texto (f) y su posición en el ranking de frecuencias (r) están relacionadas mediante la ecuación: $r \times f = k$ donde k es una constante próxima a uno que depende del texto.

Así por ejemplo para un texto de 300,000 ocurrencias, puede esperarse aproximadamente que por ejemplo si el rango 10 es la frecuencia 3000 entonces se espera que el rango 100 sea la frecuencia 300, el rango 1000 sea la frecuencia 30, el rango 10000 sea la frecuencia 3.

Opcionalmente, en el análisis de la estructura léxica, se plantea incorporar otros procedimientos, muy útiles sobre todo para el estudio comparativo de textos, como: la comparación de las distribuciones de probabilidad de textos.

-Comparación de las distribuciones de probabilidad de dos textos

Para comparar las partes del corpus, resulta conveniente que sean de tamaño parecido. Los elementos de comparación planteados son el tamaño del vocabulario de cada parte, el número de Hapax (palabras únicas) en cada parte y las frecuencias máximas de repetición de las formas gráficas (palabras o segmentos de frases).

Para medir la diferencia léxica entre dos textos se propone comparar sus distribuciones de probabilidad. El propósito es medir la diferencia absoluta entre los textos, sin considerar a ninguno de ellos como punto de referencia, se propuso la medida C_d para comparar las distribuciones. Esta medida se expresa como el

cociente del área de diferencia entre el área máxima de las distribuciones de probabilidad.

Esta medida refleja la diferencia entre los corpus y no se ve afectada por las diferencias relativas de cada uno de los términos. Para una revisión más detallada de estos conceptos véase (Montes-y-Gómez, Gelbuhh, & López-López, 2001)

El resultado final propuesto es un listado alfabético de formas y segmentos claves, acompañadas de su frecuencia absoluta y relativa con respecto al corpus total, así como de su longitud; al final del mismo se aportan también los siguientes datos estadísticos: número de formas, número de palabras (extensión del corpus), frecuencia máxima, frecuencia promedio y tasa de repetición, así como una relación del número de palabras según longitudes y frecuencias, acompañadas de sus correspondientes diagramas de barras.

B.3. Lematización y Desambiguación

Sobre el índice del corpus obtenido realizamos la corrección del corpus codificado por supresión o fusión de formas gráficas. Se puede realizar una eventual selección de formas en función de su frecuencia o longitud.

Se pueden procesar las palabras con base en diferentes criterios y propiedades, por ejemplo:

- Propiedades: Homónimos⁴⁰, polisemismo⁴¹, sinónimos⁴², hiponimia⁴³. Además, la frecuencia de aparición de las palabras se basan en una distribución de potencia (un pequeño número de palabras que se repiten mucho, y muchas palabras que se repiten muy pocas veces).

⁴⁰ Homónimos son palabras que sencillamente son similares pero tiene un significado parcial o totalmente diferente.

⁴¹ Se llama polisemia a la capacidad que tiene una sola palabra para expresar muy distintos significados. Pluralidad de significados de una palabra o de cualquier signo lingüístico y de un mensaje, con independencia de la naturaleza de los signos que lo constituyen. Al igual que la homonimia, en el caso de la polisemia se asignan varios significados a un solo significante. Pero, mientras la homonimia se produce por coincidencia de los significantes de diversos signos, la polisemia se debe a la extensión del significado de un solo significante

⁴² Sinónimos son palabras que tienen un significado similar o idéntico, pero tienen distinta escritura y ortografía, aunque se refieren a las mismas cosas

⁴³ Se denomina hipónimo a aquella palabra que posee todos los rasgos semánticos, o semas, de otra más general, su hiperónimo (término general que puede ser utilizado para referirse a la realidad nombrada por un término más particular), pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda. Por ejemplo, descapotable es hipónimo de coche, ya que comparte todos sus rasgos mínimos, a saber [+vehículo], [+con motor], [+pequeño tamaño], etcétera, pero añade a estos el rasgo [+sin capota]

- Palabras vacías: Palabras que desde un punto de vista no-lingüístico, no portan ninguna información. Lo normal es quitarlas para procesar el texto. Palabras como: a, hacia, y, para, por, cuando, etc.
- Lematización: Normalmente una misma palabra se puede representar de diferentes maneras. Esto representa un problema porque se escriben diferente pero tienen un significado similar (ej. aprender, aprendía, aprendiendo,...). Lematizar es el proceso de transformar una palabra en su raíz. Existen algoritmos⁴⁴ muy buenos para el inglés, pero para otros idiomas el estado del arte⁴⁵ no está tan avanzado. En inglés, se basa en un conjunto de reglas, en las que dado un sufijo de una palabra, lo cambian por otro, por ejemplo "TIONAL"->"TION" (ej. conditional->condition).
- Relaciones Léxicas: En inglés, existe una base de datos muy completa - desarrollada por Princeton- con 4 bases de datos (nombres, verbos, adjetivos y adverbios) llamada WordNet⁴⁶.

Cada base de datos guarda las relaciones de sinónimos entre las entradas. Las posibles relaciones que tiene son: hypernym (desayuno->comida), hyponim (comida->almuerzo), has-member (clase->alumno), member-of (copiloto->tripulación), has-part (mesa->pata), part-of (rueda->coche), antonym (arriba->abajo). Wordnet se puede bajar, o buscar en línea.

Un problema presente siempre en estos análisis es el acercamiento o alejamiento indebido entre textos causados en el primer caso por formas gráficas de construcción no eliminadas en el proceso previo, y en el segundo caso por la utilización de plurales y sinónimos. Esta problemática es tema de investigaciones actuales, y aunque ya existen avances importantes la disponibilidad de herramientas para este tipo de análisis es restringida.

La edición del diccionario completo del todo el corpus puede ayudar a detectar algunos de tales problemas. Por ejemplo formas singulares y plurales e incluso algunos errores de mecanografía pueden ser rápidamente ubicados.

La acción a tomar depende entonces de las circunstancias. Téngase en cuenta que algunas formas clave pueden resultar eliminadas de manera errónea por los defectos mencionados.

⁴⁴ <http://tartarus.org/~martin/PorterStemmer/>

⁴⁵ <http://doi.ieeecomputersociety.org/10.1109/SPIRE.2000.878189>

⁴⁶ <http://wordnet.princeton.edu/>

La Lematización y la eliminación de ambigüedades son operaciones habituales en estudios cuantitativos del vocabulario, cuando los recursos de partida se realizan manualmente.

Los lematizadores se ven obligados a definir las reglas de equivalencia entre formas y lemas y solo pueden lograr la separación automática del texto utilizando diccionarios de raíces y de sufijos; pero no existe consenso entre los lematizadores sobre cuáles son las reglas que deben seguirse, cualquier intento de lematización automática será parcial.

La eliminación automática de ambigüedades presenta más problemas, puesto que requiere la interpretación sintagmática y gramatical de la forma con el necesario retorno al texto, precodificación o examen de los contextos en que aparece.

-Lematización

La idea propuesta en la tesis de la lematización manual es realizar una rápida depuración del glosario, según el dominio de estudio y a criterio del analista juntar bajo una misma forma (o lema), varias palabras o bajo un mismo cuasisegmento varios segmentos.

Cuando la información presente en los textos ha sido filtrada y normalizada adecuadamente, se puede llegar a niveles de conocimiento bastante sofisticados.

Esta es una particularidad verdaderamente interesante del análisis y constituye la base para la utilización de la técnica en muchos problemas prácticos. Esta característica puede observarse en la experimentación del capítulo 4.

Lemas⁴⁷

Un riesgo implícito en el uso de la forma como unidad radica en que no siempre se corresponde biunívocamente con la palabra, que es una unidad básica de significación. Las variaciones de género y número, las terminaciones de los verbos y otros sufijos pueden afectar a una palabra dando lugar a que esta aparezca con varias formas (por ejemplo: alumno, alumna, alumnado).

⁴⁷ lema (todos los vocablos que cuentan con una misma raíz y con significado equivalente, es decir, una familia de palabras)

El análisis principal se plantea realizarlo a partir del conjunto de “palabras principales” (sustantivos, adjetivos, verbos, algunos adverbios), que serán a su vez sometidas a una reducción (a sus raíces), deviniendo así “morfemas lexicales”⁴⁸. Algunos autores proponen reunir bajo una misma forma, denominada “lema” a todos los vocablos que cuenten con una misma raíz y con un significado equivalente. (Etxeberria, García, & Rodríguez, 1995) Los criterios sintácticos se aplican al corpus conformado por lemas, eliminando algunos lemas dependiendo de su categoría sintáctica. Por ejemplo “desconocer”, “desconocerlos” y “desconocía” tienen la raíz léxica “desconoc”.

La unidad básica de recuento que se emplea, a posteriori se puede lematizar, es decir reagrupar las distintas inflexiones de un verbo en el infinitivo; el singular y el plural de un sustantivo en el singular; el masculino y el femenino de un adjetivo en el masculino.

De este modo, al mismo tiempo, podría ser eliminada ciertas ambigüedades que presentan formas homónimas derivadas de lemas diferentes.

Esta técnica sirve para reducir variantes morfológicas de la formas de una palabra a raíces comunes o lexemas; básicamente, esto consiste en remover el plural, el tiempo, o los atributos finales de la palabra.

Para implementar esta técnica, en el tratamiento estadístico se asigna a cada palabra y/o segmento diferente un número de orden, el mismo siempre que aparece. La numeración da lugar a la creación de un diccionario de formas propio de cada análisis. En el caso en que se procede a lematizar, las palabras que se declaran equivalentes les corresponde el mismo número. No basta sustituir una palabra por su lema o raíz, sino que se debe modificar la representación del corpus. La forma de hacerlo consiste en representar cada lema junto con el número de apariciones del mismo a lo largo del documento y no repetir el lema de la palabra en cada ocurrencia. Esta apreciación no es exclusiva para los lemas, puesto que ocurre lo mismo con las palabras.

El objetivo de esta reducción es eliminar la variabilidad de formas de una palabra, privilegiando únicamente su significado esencial. La elección de esta reducción se debe al procedimiento estadístico que sirve a la clasificación de unidades de contexto: la clasificación jerárquica descendiente que es una técnica derivada del

⁴⁸ El morfema lexical es la raíz de la palabra que remite a la familia de la cual proviene, opuesto a los morfemas derivativos que son las terminaciones que indican un matiz semántico o los morfemas flexivos que indican el género o la cantidad. Así, en la palabra “niño”, el morfema lexical es *niñ*; sus terminaciones *o*, *os*, *a*, *as*, son los morfemas flexivos, mientras que terminaciones como *ito*, *ita*, *ote*, *ota*, corresponden a los morfemas derivativos.

análisis factorial de correspondencias destinado al tratamiento de tablas binarias casi vacías, con más de 90% de ceros (Reinert, 1993) (Benzécri J. , 1981)

Esta posibilidad no se puede aplicar a todos los corpus. Por ejemplo en algunos estudios ‘tratamiento’ y ‘tratamientos’ no remitían al mismo significado. Se recomienda en este protocolo entonces dejar la lematización para una fase ulterior.

Lematizar / no lematizar es una discusión antigua que no pierde su vigencia. En cualquier caso, es indispensable reflexionar siempre sobre qué se desea contar, tomar la decisión en función del problema y de los textos. Sobre todo, no se debe lematizar demasiado pronto y a menudo es aconsejable comparar los resultados obtenidos con los tratamientos efectuados sin y con lematización. (Bécue, Análisis Estadístico de Textos, 1997).

En el caso de estudio comparativo de textos, una vez segmentado el corpus, se plantea la alternativa de sustituir, de manera automática, un “lema”, su correspondiente entrada del diccionario (infinitivo para los verbos, masculino singular para los adjetivos, etc).

En este protocolo, se invita aplicar la Lematización automática⁴⁹ para el estudio comparativo de textos, ya que en fases posteriores, será necesaria para el análisis de palabras asociadas. Para los estudios de las tablas lexicológicas se recomienda el uso del glosario editado en forma personalizada por el analista (lematización manual).

En todo caso, los algoritmos automáticos de lematización, son recomendables cuando el vocabulario, en general, esté constituido por muchas palabras distintas (más del 60%) de poca frecuencia. La gama de frecuencia nos da una idea de esto.

Se ha comprobado que esta etapa no es indispensable y puede presentar problemas (Lebart L. , Analyse statistique des réponses libres dans les enquêtes par sondage., 1986). Por ejemplo, una palabra puede estar ligada a un contexto y a un contenido particular, y ciertas preposiciones, artículos, etc. (denominados palabras herramientas), pueden caracterizar de manera efectiva las actitudes u opiniones. Por ello, es conveniente analizar bien los contextos de utilización de las palabras, antes de proceder a la lematización.

En lo sucesivo, se empleará el término forma gráfica, que podrá ser palabra o lema, según la opción escogida por el analista.

⁴⁹ Para obtener los lemas correspondientes a cada palabra, se puede utilizar la herramienta diseñada por el grupo IXA (<http://ixa.si.ehu.es>) que obtiene para cada palabra del documento, el lema que le corresponde así como la categoría morfosintáctica de la misma.

Cuasisegmento

Otra unidad compleja propuesta es el cuasisegmento (Bécue, Análisis Estadístico de Textos, 1997), que generaliza al segmento repetitivo; un casi-segmento está compuesto de varias palabras vecinas, pero no obligatoriamente continuas. Son unidades formadas por palabras que aparecen en una determinada secuencia existiendo entre ellas una distancia máxima de separación fijada, medida en números de palabras.

Por ejemplo los segmentos “formación de profesores” y “formación de los profesores” son diferentes segmentos pero ambos constituyen un cuasisegmento. Permite tomar en cuenta expresiones como “hacer deporte”, aunque se presente como “hacer un poco de deporte”, “hacer algo de deporte”, etc.

No diferenciar los homógrafos (palabras con varios significados) no suele traer consecuencias graves en cuanto al comportamiento estructural de los textos, aunque puede dificultar las interpretaciones. Se propone realizar un estudio del contexto (concordancia) de aquellas palabras que pueden tener más de un significado (según su contexto).

Por lo tanto, los resultados obtenidos son enriquecedores por la introducción de contexto (concordancias⁵⁰) en las palabras, que hace desaparecer gran parte de las ambigüedades. Aunque no es una definición formal, se puede sostener que una concordancia es el despliegue de una palabra en contexto.

-Desambiguación

La tarea de desambiguación léxica propuesta consiste en identificar la categoría gramatical de una palabra, en un contexto concreto, dentro de un conjunto determinado de categorías posibles.

La idea clave para la mejora de la desambiguación es incrementar la cantidad de información de la que se hace uso. Se hace uso de la información contenida en los textos para computar el grado de pertenencia del término a cada sentido.

Las palabras claves se han utilizado para explorar conectividad y consistencia dentro de una colección de documentos.

⁵⁰ Es una colección que recoge todas las apariciones de una palabra en un texto o conjunto de textos; en referencia al conjunto de los contextos de una cierta forma denominada forma-polo, permitiendo localizar cada una de las ocurrencias en el corpus. (Bécue, Análisis de Datos Textuales, Métodos estadísticos y algoritmos, 1991)

La lengua natural es ambigua y la misma palabra clave puede expresar significados enteramente diversos. El problema inverso es que diversas expresiones pueden referir al mismo significado, ("coche" y "automóvil"). La desambiguación semántica de tales palabras con varios significados se resuelve normalmente con el estudio del contexto o concordancia.

Hasta la fecha, los principales métodos utilizados en la desambiguación léxica abarcan desde las implementaciones en el campo de la Inteligencia Artificial hasta el empleo de técnicas estadísticas basadas en datos de corpus, pasando por el recurso de las bases de datos computacionales, como los diccionarios electrónicos. Un buen resumen de los distintos métodos es el que proporcionan Ide y Véronis (Ide & Véronis, 1998).

En esta propuesta metodológica, se ha optado por las técnicas basadas en el tratamiento cuantitativo de datos de corpus. La hipótesis es que, por lo general, toda la información necesaria para interpretar el sentido de una palabra se encuentra de un modo u otro codificada en la superficie textual.

Además, en el caso de las acepciones convencionales o estereotipadas de una palabra –es decir, todas aquellas que no proceden del uso creativo de la lengua–, los datos co-textuales necesarios para la asignación de sentido son limitados y comensurables. Por tanto, son formalizables, tipificables. (Sánchez & Almela, 2006)

A continuación, se comenta algunos procedimientos de desambiguación recomendados para el protocolo propuesto..

- Tipología textual

Lógicamente, determinados tipos de texto pueden inducir unas acepciones y reprimir la activación de otras. Por ejemplo, en un manual para el usuario de hardware informático, las probabilidades de que ratón active el sentido “dispositivo para mover el cursor sobre una pantalla” serán notablemente mayores que en un libro de texto de ciencias naturales.

Por casos como este, es que la información acerca del tipo textual puede contribuir a predecir la acepción. (Sánchez & Almela, 2006)

- Contexto sintagmático

El examen de las frecuencias lexicales prioriza el eje paradigmático de la organización textual, dando acceso – por los puntos de "densidad" – al sistema de relaciones de sustitución que le subyace.

Este enfoque debe ser entonces complementado por una reconstrucción parcial de los contextos de ocurrencia⁵¹ de las formas léxicas relevantes, pues es evidente que el sentido del discurso se construye tanto en la selección de las unidades como en su combinatoria sintagmática. Tal reconstrucción puede ser realizada eficazmente mediante la generación de concordancias. Es importante determinar el significado de una palabra en un contexto particular.

Puede ser interesante localizar cada una de las ocurrencias del documento y listar todos los contextos de una misma forma, limitándolos a una cierta dimensión en función de las necesidades particulares de cada estudio. El conjunto de los "contextos" de una cierta forma, se denomina concordancia de la forma.

Existen diversos conceptos relacionados con el estudio de las concordancias. Una noción básica que une a todos estos conceptos es la idea de la colocación o co-selección que, básicamente, se refiere a la co-ocurrencia de una forma lingüística con otras formas que la acompañan en el contexto inmediato. La concordancia o cadena de búsqueda aparece al centro, la que también recibe el nombre de nodo. Por lo tanto, una concordancia es sencillamente un listado de todas las ocurrencias de una palabra dada en un corpus acompañada de su contexto (Sinclair, 1991). La palabra buscada aparece destacada en el centro, lo cual permite analizar y detectar con rapidez sus "colocados" o palabras que aparecen en su entorno. Se llama "unidad contextual" al conjunto de contexto anterior + palabra-clave + referencia + contexto posterior.

El resultado final propuesto es un listado formas y segmentos personalizado, acompañadas de su frecuencia absoluta y relativa con respecto al corpus total. En el enfoque de estudio comparativo de textos se obtendrá un listado de lemas que se utilizarán en el estudio de especificidades.

B.4 Seleccionar las formas léxicas a conservar

Se sintetizan las principales actividades propuestas para escoger⁵² las unidades de análisis elementales que se utilizarán para los procedimientos posteriores.

⁵¹ Pensemos en el texto como una sola secuencia lineal (una frase sigue inmediatamente a la otra y así sucesivamente) en la cual cada unidad léxica ocupa una posición determinada y exclusiva. Cada unidad tiene un "contexto" propio y único: el de las unidades que se hallan a su derecha y a su izquierda. "Contextualizar" una forma léxica es observar de manera sistemática los contextos sintagmáticos de todas sus ocurrencias. (Armony, 2002)

⁵² Es indispensable reflexionar siempre sobre qué se desea contar, tomar la decisión en función del problema y de los textos. Sobre todo, no se debe lematizar demasiado pronto y a menudo es aconsejable comparar los resultados obtenidos con los tratamientos efectuados sin y con lematización. (Bécue, Análisis Estadístico de Textos, 1997)

Quizá uno de los aportes más importantes de la estrategia metodológica es la posibilidad de utilizar categorías de una manera relativamente diferente a la tradicional. Hay que recordar que cada forma gráfica presente en los análisis es tratada como una categoría de la variable léxica. Por lo que las categorías en un problema general pueden caracterizarse por formas gráficas adecuadas. Sin embargo, el analista debe ser fiel a los datos hasta donde sea posible.

Es también importante tomar en cuenta que no hay un criterio claro establecido para determinar la confiabilidad de los resultados en cuanto al número de unidades de contexto elementales incluidas en el análisis y de aquéllas eliminadas. Entre mayor sea la cantidad de texto tomada se tendrá un mejor análisis, pero sería preferible establecer un criterio objetivo de confiabilidad al respecto.

-Selección de las formas que se conservarán

En una primera etapa se sugiere contar todas las palabras, sin decidir (y por lo tanto de forma subjetiva) cuáles son importantes. La forma de proceder que permitirá desvelar lo significativo del corpus es contar las palabras contándolas a todas, sin prejuzgar nada sobre su significado, ni su contenido. (Bécue, Análisis Estadístico de Textos, 1997).

Se puede definir dos tipos de formas simples: las “palabras principales” y las “palabras relacionales”, estas últimas corresponden a aquéllas palabras que sirven a la construcción sintáctica de la frase (artículos, conjunciones, preposiciones, pronombres) y serán eliminadas del análisis principal, aunque serán tomadas en cuenta en un análisis suplementario, lo cual permite observar el papel que juegan en la formación de los segmentos repetidos.

En el conjunto de palabras que constituyen el vocabulario empleado en los textos, las frecuencias más altas suelen corresponder a vocablos que desempeñan gramaticalmente un papel auxiliar y soportan una carga semántica de segundo orden. Frente a sustantivos, adjetivos, verbos y determinados adverbios, que crean los contenidos semánticos del texto, la mayoría de las conjunciones, preposiciones, determinantes (demostrativos, artículos, posesivos, indefinidos), pronombres y otras palabras funcionales, que resultan ser en conjunto las más frecuentes, se limitan a unirlos, introducirlos, presentarlos o matizarlos.

Existe una cierta conveniencia de una intervención sobre palabras antes de someterlas a diferentes análisis. Posiblemente la supresión de los términos con un menor contenido significativo sea una de las intervenciones más habituales. Las palabras de altas frecuencias, podrían ser eliminadas con el propósito de reducir el número de palabras a estudiar y clarificar las estructuras significativas que puedan obtenerse a partir de análisis posteriores.

Se puede elaborar una lista de palabras a eliminar teniendo en cuenta la frecuencia de aparición de las mismas en el corpus.

El algoritmo para calcular un umbral mínimo de frecuencia para seleccionar las palabras se documenta en Bolasco, (Bolasco, 1999) y prevee los pasos siguientes:

- detección de la gama de las frecuencia bajas, que, a partir de la frecuencia mínima "1", es definida por el primer "salto" en los valores crecientes de las frecuencias;
- elección del valor de umbral que, según las dimensiones del corpus, corresponde al valor mínimo en el primer o en el segundo decil (10% o 20%) de la gama.

El glosario es una guía para escoger un umbral de frecuencia y/o umbral de longitud para seleccionar las formas (palabras o segmentos) que se conservaran. Para los estudios se decide conservar las palabras repetidas un número "n" de veces.

Un umbral alto diferencia los textos a partir de su estructura, tal como ésta viene indicada por las formas herramientas, en general palabras de alta frecuencia. Un umbral bajo, por su parte, permite comparar el léxico de los textos, vocabulario especializado.

En definitiva, para que el análisis estadístico tenga sentido, será necesario que las formas aparezcan con una frecuencia mínima, por ello normalmente se eliminan las formas poco frecuentes del corpus, escogiendo un umbral de frecuencias por encima del cual conservamos las formas.

Téngase en cuenta que las formas son eliminadas únicamente para los análisis estadísticos, es decir las formas por debajo del umbral de frecuencias continúan en el corpus pero desaparecen de las tablas léxica y léxica agregada.

En el conjunto de palabras críticas obtenido son de particular importancia los términos subrepresentados, es decir, aquellos términos que están presentes de forma abundante un texto y, que por el contrario, su frecuencia en el otro texto es relativamente pequeña.

Cuatro subgrupos de palabras críticas:

- Vocablos específicos del dominio: Esto es de particular cuidado ya que dentro de la aplicación dichos términos son muy comunes.
- Deícticos gramaticales: Dentro del conjunto de palabras críticas también encontramos deícticos gramaticales, es decir, vocablos que acompañados

con un gesto sirven para identificar elementos del discurso, por ejemplo ahí, aquí, éste. La mala representación de estos términos es decisiva dada la naturaleza multimodal de la aplicación en cuestión.

- Verbos involucrados en expresiones de cortesía: Las formas léxicas de verbos como: poder o querer están subrepresentados. Este resultado es de esperarse dado que las formas léxicas de estos verbos son abundantes en el habla oral espontánea y son casi totalmente ausentes en el lenguaje escrito.
- Palabras vacías: Del otro extremo del panorama encontramos un conjunto de palabras sobrerrepresentadas. Dichas palabras generalmente son artículos y preposiciones.

Entre las palabras que son más frecuentes en un texto, por el contrario, se encuentran las que se denominan "palabras vacías". En su mayor parte se trata de conjunciones, preposiciones, artículos y cópulas que no son útiles para la expresión poética o conceptual, sino sólo para la coherencia sintáctica. Al ser siempre las formas más frecuentes, estos datos no son significativos.

La información importante está transportada con las palabras funcionales, (negaciones y los adverbios). Así que quitándolas del análisis las relaciones y la cohesión entre las palabras se pierde.

Tal información es crucial en el reconocimiento de intenciones comunicativas pues capturan actos comunicativos especiales tales como preguntas, peticiones, quejas y recomendaciones, que son especialmente útiles al explorar actividades de la toma de decisión. (Bucheli, 2006)

Tradicionalmente, los investigadores pasan por alto las palabras vacías, pero hace poco tiempo algunos estudiosos han señalado la posibilidad de que, precisamente por su pobreza semántica y la facilidad de su sustitución (pensemos, por ejemplo, en las palabras del español "dentro" y "en", "a" y "hasta", etc.), podrían formar, si se las toma en su conjunto, cadenas de texto o segmentos idénticos que el individuo utiliza de manera inconsciente, que apuntarían a una posible semejanza ancestral y, por lo tanto, a una especie de huella dactilar, a un patrimonio genético del estilo de un autor del texto.

También hay palabras cuya elevada frecuencia no es estadísticamente significativa, pero que indica la presencia de motivos y temas dentro de un texto: redes isotópicas con gran importancia cohesiva aunque no se encuentren entre las apariciones escasas. Por lo tanto, es importante comprobar su presencia también en el metatexto.

Los procedimientos del estudio de la riqueza del vocabulario son de gran utilidad para seleccionar las formas, pero no hay que perder de vista el objetivo de

estudio, las formas gráficas y segmentos que finalmente se seleccionarán, también están relacionadas con el tema de interés.

Es posible, en este trabajo, apoyarse en una ingeniería lingüística⁵³, puesto que el "conocimiento objetivo" es un conocimiento escrito, es decir, expresado en el lenguaje escrito. La hipótesis en el terreno de la lingüística computacional, es que la "variación" es un fenómeno lingüístico que puede ser utilizado como indicador de conocimiento en el análisis lexicométrico de textos. A continuación, se pueden ver los tres grandes grupos de variación lingüística:

- La *variación flexional*, las formas singular y plural de los sustantivos, y las formas verbales.
- La *variación sintáctica*, se tratan tres clase de variación sintáctica
 - o La *variación de inserción*, ella concierne a toda palabra al interior de un grupo nominal
 - o La *variación de permutación*, ella implica todas las palabras o grupos de palabras que pueden permutar alrededor de un elemento (preposiciones o secuencias verbales)
 - o La *variación de coordinación*, ella concierne a todas las formas coordinadas entre palabras (sustantivos y adjetivos)
- La *variación morfo-derivacional*, integra en una terminología los fenómenos de adjetivación y sustantivación (sustantivación del adjetivo, sustantivación del verbo, adjetivación del sustantivo)

Dándole a estos fenómenos lingüísticos una formulación cuantitativa, se busca producir un nuevo tipo de indicador, un indicador lingüístico de conocimiento más rico (desde el punto de vista de la información que aporta) que las simples palabras claves que representan verdaderos estereotipos. (Cabrera Varela, 1987). El resultado final es un listado "depurado" de formas acompañadas de su frecuencia absoluta y relativa con respecto al corpus total, al final del mismo se aportan también: listado de todas las ocurrencias de una palabra ambigua dada en el corpus acompañada de su contexto. Glosario final de formas, segmentos seleccionados, con el vocabulario a estudiar acompañado de su frecuencia; al final del mismo se aportan también: listado de todas las ocurrencias de una forma clave seleccionada acompañada de su contexto.

⁵³ Se llama lingüística computacional o informática (o informática lingüística), el procesamiento informático del lenguaje natural, permitiendo a los instrumentos infométricos y definir indicadores lingüísticos. (Polanco, Transformación de la información en conocimiento y del conocimiento en decisiones estratégicas, 2002)

C. Estudio de la tabla léxica base

Los textos pueden ser numerados de manera completamente automática por el analista y, a partir del diccionario de formas, construir tablas rectangulares que serán analizadas mediante las técnicas multivariantes descriptivas, como el Análisis Factorial de Correspondencias y el Análisis de Clasificación (Benzécri, 1973) (Benzécri J. , 1981) (B & Pagès, 1992) (Lebart, Morineau, & Piron, Statistique exploratoire multidimensionnelle , 1995).

Estas tablas recogen la distribución de palabras entre los distintos textos individuales o entre grupos de individuos.

C.1. Creación de tabla léxica base

Se sintetizan las principales actividades propuestas en la tesis para crear la tabla léxica base.

Se busca destacar las semejanzas y diferencias entre unidades del contexto (subconjuntos del corpus o contextos elementales), para lo cual se propone utilizar los métodos de análisis factorial. Para aplicar los métodos de análisis multidimensional a datos textuales se plantea construir tablas de contingencia⁵⁴ particulares, en este caso la tabla correspondiente sería: Tabla léxica base.

Una vez que se ha segmentado el documento en palabras, se puede “traducir” cada una, por un número que son las repeticiones en que aparece en cada texto y ver el corpus como una sucesión de enteros.

Se puede ver entonces, el documento codificado como una tabla de doble entrada que tiene tantas filas como textos individuales y tantas columnas como palabras o segmentos del corpus. La celda de cruce de esta tabla contiene la frecuencia de repetición con la cual la palabra, segmento o casi-segmento ha sido utilizada en el texto individual.

Matriz de i filas, una por cada partición del corpus, y un número de columnas igual al de formas utilizadas por el conjunto de individuos.

En la intersección de la fila i y la columna j , figura la frecuencia con que aparece la forma j en la respuesta del individuo i

⁵⁴ las palabras o segmentos cumplen el papel de las modalidades de una variable nominal, la totalidad de las palabras contenidas en las respuestas aparecen como columnas y cada fila corresponde a una persona. En el cuerpo de la tabla aparecen las frecuencias con que cada individuo utilizó cada palabra en su respuesta libre. Todas las tablas son simétricas. Se puede trasponer las tablas y el análisis y resultados serán los mismos. Cuando el vocabulario es mucho se suele colocar en filas y los individuos en columnas.

El objetivo al construir esta tabla es comparar los perfiles léxicos de cada uno de los textos.

Resultados

Estudio comparativo de textos

- Tabla léxica - perfiles textos (textos x formas). "Tabla léxica" que contiene la frecuencia relativa con la que cada forma gráfica ha sido empleada por cada texto individual; es un tabla de contingencia que contiene los perfiles léxicos de las particiones del corpus.

C.2.Representación de tipologías

Se sintetizan las principales actividades propuestas para la construcción de representaciones de la tabla léxica base.

A estas tablas se indica aplicar el método de análisis factorial llamado análisis de correspondencias simples y los métodos de clasificación (Benzécri J. , 1981) (Lebart & Salem, 1994) para la construcción de representación de texto. Los métodos de análisis de datos no tratan de saber lo que dicen, pero si de saber, sí dicen lo mismo.

-Análisis factorial de correspondencias

Como todos los métodos de análisis factorial, el análisis de correspondencias permite la extracción de nuevas variables - los factores - que resumen de una manera organizada la información significativa contenida en los innumerables datos de las tablas; además, esta técnica de análisis permite la creación de gráficos que muestran - en uno o más espacios - los puntos que identifican los objetos en filas y/o en columnas, que - en nuestro caso - son las entidades lingüísticas (palabras, lemas, segmentos de textos y textos) con sus respectivas características de proveniencia.

En una tabla de contingencia, las filas y las columnas representan dos particiones de una misma población y ambas particiones juegan un papel análogo: para analizar el contenido de la tabla tiene sentido considerar tanto la nube de puntos-fila como la nube de puntos- columna.

El análisis de correspondencias ofrece una representación gráfica conjunta de ambas; para ello efectúa la proyección de las nubes sobre subespacios de dimensión reducida pero manteniendo la máxima dispersión posible. Proporciona una visión gráfica, simplificada, de la información que contiene y pone en evidencia las diferencias entre los distintos perfiles léxicos. Permitirá ver qué palabras y/o qué expresiones diferencian a los individuos.

El análisis de correspondencias es muy sensible a las rarezas de los textos, lo que permite poner de relieve las partes del corpus cuya distribución del vocabulario es muy peculiar (segmentos de texto situados a la periférica, alejados de los otros). Por otra parte, pone en evidencia las oposiciones globales, en las cuales entran en juego el conjunto de palabras y partes; lo que permite visualizar retrocesos y las similitudes entre partes, aunque estén separadas temporalmente.

Si en una investigación se busca establecer las diferentes formas de expresión frente a un determinado tema en una población relativamente heterogénea, este tipo de análisis puede resultar muy provechoso.

Por otro lado, si la población a ser interrogada es muy homogénea y la temática abordada es especializada y de dominio de la población, se debe tener en cuenta que es normal llegar a la obtención de un gran grupo en donde se ubican la mayoría de las personas y formas gráficas, y la aparición de otros pequeños grupos que se distinguen más por algunos matices especiales en su respuesta que por pensar realmente de una manera diferente. Nótese que para una pregunta no especializada, como puede ser el caso de un tema político, el análisis puede entregar mejores resultados.

Si se observa la inercia de cada respuesta o texto individual, y se ordena de mayor a menor según la misma, se puede determinar que particiones del corpus son las más dispersas o extremadamente diferentes (en su vocabulario). Cuando se agrupe en cluster, cada uno de estos textos formará un grupo homogéneo pero con un único individuo.

Después de hacer un primer análisis de correspondencia, se podrá seleccionar (basado en la inercia), cuales son estos “textos extremos” o especializados y proyectarlos de manera ilustrativa en un segundo análisis, para posteriormente clasificarlos en una partición según su proyección en el plano factorial. De esta manera no influyen en la distribución del resto de los textos, permitiendo ver con claridad el resto de las relaciones. Más adelante en el estudio de especificidades se podrán estudiar el vocabulario exclusivo de estos textos.

Se recomienda hacer este análisis de la inercia de los “textos o respuestas individuales” siempre que en el plano se visualicen estos “textos extremos” y/o en una posterior agrupamiento, luego de armar el árbol de agregación o dendrograma, se observa en las particiones de pocos grupos, cluster de un único texto.

En general, los segmentos suelen ser proyectados de forma ilustrativa en los planos factoriales para elaborar explicaciones, sin formar parte de la nube de puntos, siendo estas unidades derivadas de otras más simples. En el análisis el contexto de las formas es ignorado por completo en las primeras fases. La proyección de segmentos repetidos, el cálculo y edición de respuestas

características, la proyección de variables nominales permiten por un lado la recuperación de los contextos de las formas y por otro lado la recuperación de los discursos con toda su carga emotiva.

Debe tenerse en cuenta de todas maneras que en la base del método, existen sesgos debidos a la falta de contextualización. De acuerdo a la investigación que se realice y a la disponibilidad de tiempo y recursos se puede intentar mejorar los análisis, mediante la contextualización de ciertas formas clave.

-Clustering sobre factores

El Análisis de Correspondencias compara la información conocida sobre los elementos de un conjunto y visualiza los resultados de esta comparación. La Clasificación pretende resumir esta información; para ello, realiza un trabajo de síntesis que consiste en reagrupar los elementos en un número pequeño de clases homogéneas respecto a la información, obteniendo una partición tal que la clases sean homogéneas y bien separadas, es decir que las distancias entre individuos de una misma clase o grupo sea pequeña y la distancia entre distintas clases sea grande.

La aplicación de los métodos clásicos del análisis estadístico descriptivo se ha limitado a simplificar la información, o sea darle una dimensión reducida para hacerla más rápidamente accesible, sin embargo el Lenguaje se resiste a esos tratamientos demasiados simplificadores. Es por ello que el análisis textual por el análisis descriptivo multidimensional, promete una aplicación más pertinente.

Se muestra en este protocolo una propuesta de agrupamiento de corpus textuales mediante la concatenación de métodos de agrupamiento ya existentes.

En razón que los individuos no se expresan de la misma forma (según su pertenencia a un grupo socioeconómico, su edad, su nivel de educación, sus opiniones, etc.), parece tener sentido plantear la necesidad de agruparlos según su vocabulario, para después, caracterizar las clases así obtenidas por la información conocida sobre los mismos.

Al agrupar textos individuales es muy útil obtener simultáneamente los grupos de individuos y las palabras claves de cada grupo. Así, se divide la colección de textos en categorías más significativas y se genera una descripción compacta de cada cluster en términos no sólo de los valores de los atributos, sino también de su relevancia (Sperber, 1986).

En esta tesis se propone una variante de agrupamiento que permite simultáneamente agrupar los documentos y obtener las palabras claves que caracterizan cada grupo, sin necesidad de tener conocimiento previo del dominio.

Esta clasificación se indica realizarla por comparación de los perfiles léxicos de los individuos obtenidos del análisis de correspondencias directo de sus textos. Esta técnica permite relacionar las formas más mencionadas por los distintos individuos, es decir constituir perfiles léxicos que servirán de base para la construcción de tipologías.

Como sugiere todo enfoque tipológico, no resultan de tanto interés los perfiles aislados, serán las diferencias entre perfiles las que llamarán la atención. La propuesta es que aplicando la técnica de clasificación jerárquica sobre las coordenadas factoriales obtenidas mediante el análisis de correspondencias, se llega a la formación de grupos de individuos con perfiles léxicos semejantes.

-Proyección ilustrativa de cluster

El Análisis Factorial de la tabla de individuo*formas permite visualizar las distancias entre los perfiles léxicos de los individuos. Al proyectar las modalidades de las variables nominales sobre los ejes factoriales, se ilustran los mismos mediante una información que no ha participado en la construcción de los planos, lo que tiene gran importancia en la interpretación de los resultados. Las variables nominales están en el mismo espacio que la variable léxica y es legítimo interpretar proximidades entre formas y modalidades de variables nominales. Una forma léxica será próxima a las modalidades escogidas por los individuos que han empleado esta forma.

Los resultados propuestos del análisis se resumen a través de gráficos bidimensionales (tipo planos cartesianos) que permiten evaluar las relaciones de proximidad/distancia - o sea de semejanza/diferencia - entre los objetos considerados. Se proporcionan las medidas - en concreto los valores test - que facilitan la interpretación de los polos factoriales que organizan las diferencias/semejanzas entre los objetos considerados. Y otros indicadores como la inercia de cada punto. También se incluye Dendograma e información sobre los cluster.

C.3. Estudio de las especificidades

Las palabras juegan un rol importante para aquel que desea adentrarse en una disciplina. Conocer los términos que se utilizan en una especialidad permite acceder a un conjunto de conocimientos específicos (categorías, procedimientos o herramientas) que caracterizan y definen a una comunidad discursiva determinada, asociada a esa especialidad o disciplina. Los segmentos de frase clave son un tipo especialmente útil de información abreviada.

Se sintetizan las principales actividades propuestas en la tesis para el estudio de especificidades y caracterización de clases o tipologías. Proceso que se propone para comprobar cuáles son las unidades lexicales (palabras, lemas o categorías)

típicas o exclusivas de un texto, o de un subconjunto del corpus definido por una variable cualquiera.

-Formas, segmentos y respuestas características

El método de las especificidades de unidades lexicales típicas, consiste en determinar elementos que resultan característicos de un determinado texto cuando se lleva a cabo un estudio comparativo de varios de ellos. El carácter específico de una palabra o un segmento para un texto concreto se determina por comparación con un corpus más amplio de textos que sirve como referencia. (Etxeberría, García, & Rodríguez, 1995).

El estudio de varios textos reunidos en un mismo corpus, con el ánimo de encontrar diferencias en los mismos, lleva a confrontar la frecuencia con que aparece en los distintos textos una palabra o un segmento. Determinadas formas son muy utilizadas en ciertos textos, mientras que prácticamente están ausentes de otros. Lo que se pretende con este método de las “especificidades” es detectar las palabras que destacan por su reiterada aparición, o por el contrario, por su rareza.

Las palabras o segmentos que presentan una especificidad positiva dentro de una parte diferencial del corpus son las que se emplean por encima de lo que cabría esperar si las apariciones de esta se distribuyeran aleatoriamente en todo el corpus, por el contrario las negativas corresponden a las formas que están infrautilizadas en relación a su presencia en el corpus.

Se plantea identificar las palabras “sobrerrepresentadas” características en un texto, en comparación con la totalidad. Es decir, comparar la frecuencia relativa de cada una de las palabras en un texto, y en la totalidad del corpus, considerada como frecuencia de referencia, utilizando un test clásico para emitir el juicio comparativo.

Para facilitar la lectura de los resultados del test, se traduce la probabilidad asociada a la comparación en un valor de test, estandarizado de tal forma que se pueda leer como una realización de una variable de Laplace Gauss centrada y reducida.. Por lo tanto se puede considerar como palabras características aquellas cuyo valor de test supere el umbral de 1.96 (palabras anormalmente frecuentes), por debajo de este valor no hay significación estadística (palabras anormalmente poco frecuentes). (Bécue, Análisis Estadístico de Textos, 1997). El valor de test constituye una medida propuesta para determinar la diferencia entre la frecuencia de la palabra en el grupo y la frecuencia de esta misma palabra en el conjunto.

Al extraer las palabras características de cada texto o grupo, se ignora totalmente los contextos de las palabras. Por esto, se busca identificar los enunciados que se puedan considerar característicos (modales) de cada grupo.

Los enunciados característicos son particiones íntegras del corpus. Una sola frase nunca resumirá la riqueza de un texto como tampoco nunca un único individuo modal será un buen representante de toda una clase de individuos.

Dado un grupo de individuos (textos), se puede calcular su perfil léxico medio. Se consideran como enunciados modales de este grupo, enunciados más próximos a dicho perfil medio, según la distancia de Chi². La caracterización anterior puede mejorarse dividiendo la distancia frase - grupo por la media aritmética de las distancias de esta frase a todos los otros grupos. Se suele también seleccionar a las respuestas características siguiendo otro criterio, el criterio del valor de test medio. (Lebart, Salem, & Bécue, 2000). Para cada uno de los grupos, se afecta a cada palabra un valor de test que valora su frecuencia en el grupo comparada a su frecuencia en la muestra. Se puede atribuir a cada frase la media de los valores de test de las palabras que la componen. Los enunciados con valor medio más alto serán los más característicos del grupo.

Las unidades lexicales "típicas" se definen por exceso (sobre utilización) o por defecto (sub utilización), según un criterio estadístico que requiere la aplicación del test de chi cuadrado. Las unidades lexicales "exclusivas" son los presentes solamente dentro del subconjunto considerado y "no" en otros. En este caso, el test de Chi cuadrado contempla el repetido cruce entre dos filas (texto y corpus) y dos columnas (presencias y ausencias de cada palabra); de modo que el test tenga solamente un grado de libertad y un umbral (5%) igual al valor 3,84.

-Asociaciones de Palabras para el estudio comparativo de textos

Se propone obtener un grupo restringido de formas significativas que pueden ser consideradas co-ocurrentes (derecha o izquierda) del «polo» analizado. A cada una de ellas le corresponde un índice conforme al cual pueden ser ordenadas bien jerárquicamente, bien describiendo una red de co-ocurrencias lexicales, cuya representación consistirá en un grafo de la red lexical del «polo» estudiado.

Cada palabra tiene un sistema de "direcciones" donde se registran los lugares del corpus en la cual "funciona". Estos "lugares" son los contextos elementales, es decir, los segmentos de texto que corresponden aproximadamente a los enunciados.

Así pues, las co-ocurrencias, son las cantidades que resultan del cómputo del número de veces que dos o más elementos lingüísticos (palabras o lemas) "coexisten", es decir están presentes contemporáneamente, en los mismos contextos elementales.

El cómputo de la co-ocurrencia se hace construyendo tablas cuadradas, o sea con los mismos elementos en filas y en columnas. En términos técnicos, estas tablas se llaman simétricas, ya que las dos partes separadas por la diagonal contienen,

especularmente, los mismos valores. El estudio de las relaciones entre co-ocurrencias, se realiza a través de índices de asociación específicos.

La determinación y medida de las relaciones entre palabras se pueden lograr mediante diferentes métodos. Uno de ellos considera el entorno de cada palabra. Este entorno está integrado por "n" palabras a cada lado de la seleccionada, atribuyéndoseles diferente valor de acuerdo a su proximidad. El conjunto de relaciones cuantificadas permite atribuir a cada palabra un factor de capacidad de relación.

Después de realizar la lematización del corpus, las asociaciones entre palabras hace aparecer una estructura de red, en donde un grupo puede representarse como un grafo conexo entre las palabras más fuertemente asociadas que se convierte en una unidad por eliminación de los vínculos a otras palabras con un menor coeficiente de asociación. El criterio para que una palabra pertenezca a un grupo es que su vínculo a otra palabra sea superior a un cierto umbral o que se acepte hasta un cierto número de palabras.

Para el cálculo del índice de la asociación, puede utilizarse el coeficiente del coseno.

$$C(X,Y) = \frac{X \cap Y}{\sqrt{X} \times \sqrt{Y}}$$

Su fórmula es la siguiente:

Donde el coeficiente del coseno entre cada par de palabras (X, Y) se define como el cociente entre su cantidad de co-ocurrencias y la que se obtiene multiplicando las raíces cuadradas de las respectivas ocurrencias. En análisis de textos, el uso de esta fórmula ha sido propuesto por G. Salton (Salton, 1989).

En un gráfico, la forma seleccionada está en el centro. Las otras se distribuyen alrededor de ella, cada una a una distancia proporcional a su grado de asociación. Por tanto, las relaciones significativas son del tipo uno a uno, entre la forma central y cada uno de las otras.

La información complementaria a este gráfico se presenta con una tabla donde:

- una columna muestra las palabras asociados a la forma "central";
- los coeficientes del coseno;
- ocurrencias de cada palabra asociada;
- co-ocurrencias de las formas asociadas y la palabra clave central..

Resultados propuestos. Para el estudio comparativo de textos se tendrá vocabulario característico de tipologías de textos o partes del corpus y un estudio de especificidades de cada texto individual (lemas típicos y exclusivos). También

se podrá obtener de manera opcional el análisis de asociación de palabras claves del corpus.

D. Estudio de la tabla léxica agregada

Si se quiere comparar “partes” del documento llamados textos con los perfiles léxicos, la tabla que lo permite es una “tabla léxica agregada”, parecida a la tabla léxica, que contiene las frecuencias de las formas en cada parte.

D.1. Creación de tabla léxica Agregada

Se sintetizan las principales actividades propuestas en la tesis para la creación de Tabla de contingencia de “individuos * grupos”.

Para aplicar los métodos de análisis multidimensional se indica construir, en este caso, la tabla léxica agregada.

La tabla léxica agregada se construye cuando el corpus es particionado en textos que se desean comparar. El propósito al construir la tabla es comparar los perfiles léxicos de los textos en los cuales se particiona el corpus. Esta partición del documento se hace según grupos de clasificación; la tabla contiene tantas filas como formas y tantas columnas como modalidades de clasificación (tipologías de los grupos).

Esta categoría de partes del corpus (respuestas o textos) es aportada por los datos complementarios cualitativos

La construcción de la tabla léxica agregada puede hacerse a partir de la tabla léxica original, haciendo uso de la matriz X, de Individuos*Modalidades.

Para el estudio comparativo de texto, se pueden utilizar la tabla léxica base “transpuesta”, es decir que en columnas tendremos a cada texto no agrupado y en fila todo el vocabulario.

Esta tabla permitirá agrupar el vocabulario según como se distribuye en cada uno de los textos. No siempre será necesario realizarlo, dependerá de los objetivos del estudio. Pero antes de analizar la tabla léxica agregada se recomienda primero analizar esta tabla teniendo en cuenta los textos individuales que se seleccionan para proyectarlos ilustrativamente. Esta tabla es simétrica a la anterior y el Análisis Factorial de Correspondencias simples es el mismo, la idea en este caso es clasificar el vocabulario, no los textos, pero siempre sobre los factores.

Resultados propuestos Tabla léxica agregada – perfiles grupos (formas x grupos textos). Cuando existen una o varias particiones pertinentes del corpus, se puede construir, para cada una de ellas, la “tabla léxica agregada” de contingencia que

contiene para cada parte, la frecuencia con la que se encuentra una forma en esta parte.

D.2. Representación de textos

Se sintetizan las principales actividades propuestas para la creación de la representación de las tablas léxicas agregadas.

- Clustering sobre análisis tabla agregada (Agrupamiento del vocabulario según distribución en grupos de textos)
- Análisis factorial de correspondencia de la tabla léxica agregada.
- Proyección ilustrativa de⁵⁵ cluster. (visualización de formas y cluster).

Se plantea analizar la tabla léxica agregada para estudiar las posiciones relativas de diferentes grupos de individuos (partes del corpus) en relación al vocabulario empleado y caracterizarlos por las palabras utilizadas. Si la partición del corpus corresponde a períodos de tiempo distintos, se podrá estudiar la evolución en el tiempo.

D.3. Estudio de las especificidades datos complementarios

Se sintetizan las principales actividades propuestas para el estudio de especificidades y caracterización de clases o grupos de textos.

Estudio comparativo de textos:

- Textos característicos, para cada cluster
- Vocabulario característico para cada grupo de textos (vocabulario de datos complementarios)
- Asociaciones de Palabras de grupos de textos (previa lematización)
- Descripción de cluster
- Descripción de datos complementarios

Se indica en el procedimiento completar la representación mediante el estudio de las características léxicas (palabras y respuestas) de las partes del corpus, en particular aquellas que ocupan un lugar señalado.

El estudio de especificidades se plantea para comprobar cuáles son las unidades lexicales (palabras, lemas) típicas de un texto, o de un subconjunto del corpus definido por una variable cualquiera.

⁵⁵ También se pueden proyectar otras variables cualitativas o numéricas.

Se propone describir y validar los cluster obtenidos en el procedimiento anterior. Se tiene grupos de formas caracterizadas por grupos de textos. Luego, tomando los datos complementarios se procede a describir las modalidades (categorías de textos), identificando formas típicas.

-Asociaciones de Palabras para el estudio comparativo de textos

Un procedimiento opcional propuesto, es agrupar los textos en función de datos complementarios y la asociación de palabras en cada grupo de textos.

Se toma cada subcorpus, es decir grupos de textos en función de variables cualitativas y se realiza la asociación de las principales palabras de cada grupo de textos. Por ejemplo, se analiza fecha por fecha, o autor por autor.

Resultados propuestos. Para el estudio comparativo de textos se tendrá vocabulario característico de tipologías de vocabulario utilizados por un grupo de textos o partes del corpus y un estudio de especificidades de cada grupo. También se podrá obtener de manera opcional el análisis de asociación de palabras claves de partes del corpus.

4. CASO PRÁCTICO

Siguiendo las estrategias expuestas para el estudio comparativo de textos, cuyo procedimiento se explica en la sección 3, se procede en los siguientes apartados a desarrollar el estudio de análisis de textos que permitirán experimentar con el protocolo propuesto.

Caso de Estudio comparativo de textos (análisis del discurso)

Estudio de la colección las réplicas y contrarréplicas entre José Rafael Sañudo y Sergio Elías Ortiz, suscitadas por la publicación en 1925 de “Estudios sobre la vida de Bolívar” que en forma de artículos divulgó la revista “Ilustración Nariñense” entre 1925 y 1928.

Los ESTUDIOS SOBRE LA VIDA DE BOLIVAR, tuvieron su aparición en Pasto en octubre de 1925. Hasta 1949 esta polémica obra cuenta con tres ediciones, (la de 1931, por Imprenta Minerva, y la de 1949, publicada por Editorial Cervantes). Guido Enríquez Ruiz, informa que “En 1925 el abogado pastuso José Rafael Sañudo publicó, editado en la imprenta de Díaz del Castillo y cía., un libro intitulado “Estudios sobre la vida de Bolívar” que el gobierno conservador de Pedro Nel Ospina se apresuró a recoger para destruirlo. ¿Por qué? Simplemente porque el bien documentado libro de Sañudo decía la verdad acerca del Libertador apartándose de la visión romántica seguida por la historiografía oficial que lo hacía un personaje inasequible, lejano a toda realidad humana y lo convertía en un superhéroe con todas las virtudes y ningún defecto.” (Enríquez Ruiz, 2009). También hay que recordar que estos ESTUDIOS, por orden del dictador Juan Vicente Gómez fueron proscritos en Venezuela. De esta obra se conoce, además, una publicación impresa por la Editorial Bedout de octubre de 1975, la cual será el referente de este trabajo, y una última, en la colección Lista negra de Editorial Planeta, publicada en 1995.

Con relación a las réplicas y contrarréplicas, en la polémica originada por la publicación en 1925 de “Estudios de la vida de Bolívar”, se han elegido las cruzadas entre Sergio Elías Ortiz y José Rafael Sañudo, en la Revista “Ilustración Nariñense”⁵⁶ entre 1925 y 1928.

⁵⁶ Wikipedia informa que Ilustración Nariñense fue un periódico colombiano de la ciudad San Juan de Pasto, el cual registraba a partir de su fundación en 1924 por Rafael Delgado Ch. (director propietario de la revista), "mensual o quincenalmente, la actividad social y económica gubernamental y de todo orden de la ciudad de Pasto; su director manejaba el primer fotograbado llegado a Pasto a la Imprenta Departamental, lo que le permitía registrar gráficamente". La revista fue impresa en los talleres del Editorial Cervantes en Pasto y era "registrada para el curso libre de porte para el servicio interior", con licencia N°525 del 18 de julio de 1937, Imprenta Imperial en Pasto (años 1920 adelante).

- La primera réplica lleva por título: *HISTORIA?*, publicada en la página 3 y siguientes del N° 9 de Ilustración Nariñense de 1925, de autoría de Sergio Elías Ortiz, quien firma bajo el seudónimo “Un indio patriota”
- La segunda réplica de Sergio Elías Ortiz, que lleva el título de *AÚN VIVE LA HISTORIA*, fue publicada en el N° 10, de la revista mencionada, en 1925, bajo el seudónimo de Alexo Corfú. En el mismo número publica, Sergio Elías Ortiz, su tercera réplica, con el seudónimo de “Un indio patriota”; este artículo lleva también el título *HISTORIA?*
- A su vez José Rafael Sañudo, responde con el artículo: “*Los refutadores del señor José Rafael Sañudo*”, publicado en Ilustración Nariñense en 1925 bajo el seudónimo de Próspero Gallo. Las otras contrarréplicas están firmadas bajo el seudónimo de Patricio Leal y Franco: *UNAS VERDADES (1)*, *UNAS VERDADES (2)*, *UNAS VERDADES (3)*, *UNAS VERDADES (4)*,

4.1 PROCEDIMIENTO

A. Preparación del corpus

Se digitaron los artículos señalados de la Revista “Ilustración Nariñense”. Con cada artículo se incluyó:

- Carácter del artículo (réplica o contrarréplica),
- Título dado al artículo,
- Fecha
- Contexto del artículo
- Seudónimo utilizado para firmar el artículo y su correspondiente autor.

A.1. Edición del corpus

Se procedió a realizar una serie de transformaciones del archivo que se está analizando: eliminación de los espacios vacíos en exceso, adición del espacio después de signos de puntuación, eliminación de comilla simple, mayúsculas de inicio de oración se pasan a minúscula y verificación de la homogeneidad ortográfica.

Usando el procesador de texto de Microsoft Word se puede usar una sencilla “macro”⁵⁷, para reemplazar la primera letra de cada oración por su minúscula, siempre que esta no corresponda a un nombre propio.

⁵⁷ A continuación se presenta la macro realizada en XXXX Sub minuscula()

x = 0

n = Selection.Range.Sentences.Count

Do

x = x + 1

Selection.Sentences(x).Characters(1).Case = wdLowerCase

A.2. Partición del corpus

Cada artículo, en una fecha dada y con un autor en particular, constituye la mínima observación a analizar (o "individuo"). Es decir que constituye una partición o "texto individual" del corpus, que será necesario identificar.

En una planilla de cálculo de Microsoft Excel, se copió de cada artículo el título, seudónimo, autor y fecha, con estos datos se armó una tabla de doble entrada donde cada fila constituye una partición del corpus (artículo) A esta tabla se le agregó otra columna con un código que identifica unívocamente a cada partición. Sabiendo que algunos corresponden a "réplicas" y otros "contrarréplicas", se optó por incluir esta diferenciación en el identificador, en el primer caso el código comienza con "DCr##", en el segundo caso "DCc##", podemos verlo en la siguiente Tabla 1.

Loop Until x = n
End Sub

Tabla 1. Artículos de la polémica entre Sergio Elías Ortiz y José Rafael Sañudo publicados en la revista "Ilustración Nariñense"

ORDEN DE APARICIÓN DE ARTÍCULO	ID	TÍTULO	FEC HA	SEUDÓNIMO	AUTOR	CARÁCTER
1	DCr_SEO_IP1	<i>HISTORIA?</i>	1925	Un indio patriota	Sergio Elías Ortiz	réplica
2	DCr_SEO_AC	<i>AÚN VIVE LA HISTORIA</i>	1925	Alexo Corfú	Sergio Elías Ortiz	réplica
3	DCr_SEO_IP2	<i>HISTORIA?</i>	1925	Un indio patriota	Sergio Elías Ortiz	réplica
4	DCc_JRS_PG	<i>“Los refutadores del señor José Rafael Sañudo”</i>	1925	Próspero Gallo	José Rafael Sañudo	contrarréplica
5	DCc_JRS_PLF1	<i>UNAS VERDADES (1)</i>	1925 (?)	Patricio Leal Franco y	José Rafael Sañudo	contrarréplica
6	DCc_JRS_PLF2	<i>UNAS VERDADES (2)</i>	1928	Patricio Leal Franco y	José Rafael Sañudo	contrarréplica
7	DCc_JRS_PLF3	<i>UNAS VERDADES (3)</i>	1928	Patricio Leal Franco y	José Rafael Sañudo	contrarréplica
8	DCc_JRS_PLF4	<i>UNAS VERDADES (4)</i>	1928	Patricio Leal Franco y	José Rafael Sañudo	contrarréplica

En este estudio, el criterio de clasificación con el cuál se hace la partición es el carácter del discurso (réplica o contrarréplica).

Habiendo identificado cada texto individual, se incluyó esta información en el archivo con el corpus. De esta manera, se reemplaza el “autor-seudónimo-carácter” por el código de identificación

A.3. Preparación de la base de datos

La planilla creada en el paso anterior, incluye variables cualitativas nominales (Tabla 1), derivadas de la información que se obtuvo junto a los discursos:

- a) Una variable de dos modalidades según el tipo de discurso, es decir si corresponde a las réplica o contrarréplica
- b) Una variable que identifica al autor
- c) Una variable que indica el pseudónimo utilizado, en el ID, se agregó un número que indica la secuencia de aparición de un mismo pseudónimo en el caso de repetición

Esto permitirá más adelante analizar las frases y formas características para cada “texto individual” (además de los textos agrupados según autor, carácter, peculiaridades de pseudónimo).

Con las variables identificadas anteriormente se preparó un fichero con datos complementarios a cada discurso⁵⁸. En la Ilustración 1., se puede ver un fragmento del fichero en formato ASCII, donde las columnas están delimitadas, en este caso, por punto y coma (dependiendo de la configuración del punto decimal del sistema operativo).

```
ID;TÍTULO;FECHA;SEUDÓNIMO;AUTOR;CARÁCTER
DCr_SEO_IP1;HISTORIA?;1925;Un indio patriota;Sergio Elías Ortiz;réplica
DCr_SEO_AC;AÚN VIVE LA HISTORIA;1925;Alexo Corfú;Sergio Elías Ortiz;réplica
DCr_SEO_IP2;HISTORIA?;1925;Un indio patriota;Sergio Elías Ortiz;réplica
DCc_JRS_PG;“Los refutadores del señor José Rafael Sañudo”;1925;Próspero Gallo;José Rafael Sañudo;contrarréplica
DCc_JRS_PLF1;UNAS VERDADES (1);1925(?);Patricio Leal y Franco;José Rafael Sañudo;contrarréplica
DCc_JRS_PLF2;UNAS VERDADES (2);1928;Patricio Leal y Franco;José Rafael Sañudo;contrarréplica
DCc_JRS_PLF3;UNAS VERDADES (3);1928;Patricio Leal y Franco;José Rafael Sañudo;contrarréplica
DCc_JRS_PLF4;UNAS VERDADES (4);1928;Patricio Leal y Franco;José Rafael Sañudo;contrarréplica
```

Ilustración 1 Fichero de datos delimitado por “;”

B. Segmentación

B.1. Elaboración de documentos lexicométricos

⁵⁸ En este caso, se guardó la planilla de Excel con la tabla de datos cualitativos con el formato “csv, delimitado por coma”

En este procedimiento se caracterizó una serie de cadenas reconocidas como nombres propios (de persona y lugar); convirtiendo las secuencias de formas gráficas reconocidas como “poliformes” o “segmentos repetidos” (Ilustración 2) en cadenas unitarias, para utilizarlas como segmentos de frase durante el proceso de análisis (por ejemplo "en otras palabras" y "en tal caso" se transforman respectivamente en "en_otras_palabras" y "en_tal_caso").

Para la desambiguación⁵⁹, se reconoce y distinguen tres clases de objetos lingüísticos como segmentos de frases repetidos: los nombres propios (de persona o lugar); las locuciones (palabras compuestas y modismos); y los tiempos compuestos.

Listado de segmentos:

El_Libertador	Volver_a_poner	Indio_patriota
República_de_Colombia	En_tela_de_juicio	Estados_y_naciones
Sentido_común	Ilustración_Nariñense	Expiación_de_una_madre
Condición_humana	José_Rafael_Sañudo	El_bien_y_el_mal
Carne_y_hueso	El_doctor_Sañudo	Boletín_de_Estudios_Históricos
Sistemas_de_ideas	Sergio_Elías_Ortiz	20_de_julio
Proyecto_político	Leopoldo_López_Alvarez	Concejo_de_Pasto
Público_y_privado	Santo_Tomás_de_Aquino	Eliseo_Gómez_Jurado
Mejor_vida	El_Dr._Ortiz	Alejandro_Santander
Social_y_político	Alejandro_de_Hales	Entique_Otero_D'Acosta
Tiempo_perdido	Rogerio_Bacon	Olegario_Medina
Tiempos_que_corren	San_Buenaventura	Filosofía_del_Derecho
Últimos_tiempos	Duns_Scotto	

Ilustración 2 Ejemplos de segmentos repetidos o multipalabras

Identificadas las palabras y segmentos repetidos, se los contó y enumeró. En la Tabla 2 se muestra el índice del corpus ordenado por la frecuencia de ocurrencia.

⁵⁹ Resolver casos de ambigüedad semántica, concretamente los atribuibles a los homógrafos, es decir, palabras con la misma forma gráfica pero con diversos significados

Tabla 2 Índice de corpus, ordenado por frecuencia absoluta (fragmento)

Orden (mayor a menor frecuencia)	Frecuencia absoluta	forma
1	1448	de
2	988	que
3	773	la
4	709	y
5	644	el
6	561	en
7	392	a
8	375	los
9	274	no
10	253	se
11	243	del
12	199	es
13	199	por
14	191	su
15	177	un
16	164	lo
17	150	con
18	144	las
19	139	como
20	126	al
21	122	una
22	120	para
23	103	Sañudo
24	94	o
25	94	sus
26	87	más
27	81	porque
28	68	le
29	66	señor
30	65	doctor
31	57	Bolívar
32	57	si
33	53	fue
34	49	cuando
35	49	sin
36	47	ser
37	43	era
38	41	ha
39	41	pues
40	40	sino
41	40	sobre
42	39	En
43	37	bien
44	36	El
45	36	ni
46	36	Y
47	35	así

48	35	dice
49	35	hasta
50	35	historia
51	34	Estudios
52	34	muy
53	34	pero
54	34	ya
55	33	La
56	33	mismo
57	33	vida
58	32	les
59	31	decir
60	31	esa
61	31	son
62	30	Pasto
63	29	ese
64	29	hay
65	28	solo
66	28	todo
67	28	uno
68	27	había
69	26	No
70	26	página
71	25	esta
72	25	expiación

Se obtuvo así una primera información sobre el corpus, la repartición de las palabras, además de su longitud y el tamaño de su vocabulario (glosario de formas gráficas):

5402 formas simples, de un total de 21646 ocurrencias
1858 formas con al menos 2 frecuencias, que corresponden a 18103 ocurrencias (88.44% del corpus)
3543 formas de frecuencia 1 (harpas) (que corresponde a 11.56% del corpus sin lematizar)

Entre las palabras claves del dominio (con significado), las más frecuentes son: “Bolívar”, “Sañudo”, “doctor”, “historia”, “estudios”, “vida”, “Pasto”.

B.2. Análisis de la estructura léxica del vocabulario

Con base en el Índice del Corpus (Tabla 2), se efectuó el estudio de la gama de frecuencias de formas. Resultados visualizados en la (Tabla 3).

Tabla 3 Gama de frecuencia de formas (fragmento)

Nº	Frecuencia	%
1	3544	16%
2	803	3,7%
3	340	1,5%
4	193	0,8%
5	78	0,3%
6	99	0,4%
7	52	0,2%
8	31	0,1%
9	27	0,1%
10	26	0,1%
11	20	0,09%
12	20	0,09%
13	18	0,08%
14	13	0,06%
15	3	0,01%
16	19	0,08%
17	8	0,03%
18	4	0,01%
19	5	0,02%
20	7	0,03%
21	9	0,04%
22	2	0,009
23	4	0,01%
24	4	0,01%
25	3	0,01%
26	2	0,009
27	1	0,005
28	3	0,01%
29	2	0,009
30	1	0,005
31	3	0,01%
32	1	0,005
33	3	0,01%
34	4	0,01%
35	4	0,01%
36	3	0,01%

Para depurar el Índice del Corpus se procedió a analizar, mediante el estudio de concordancias, las mismas formas con distinto significado (para desambiguación) y distintas formas con mismo significado (para lematización). El estudio de concordancias es un análisis muy extenso. En el glosario será necesario analizar el contexto de todas aquellas formas que el analista considere conveniente. Siendo este caso una aproximación metodológica solo se muestra en la ilustración 3 parte del estudio de la concordancia de las formas “Bolívar” y “justicia”, palabras que puede tener distinto significado según su contexto.

Para las interpretaciones posteriores en los estudios de las tablas, será necesario el análisis de las concordancias. En este trabajo aunque se realizaron los análisis no se presentan todos los resultados.

54	a orden es también una verdad histórica, donde no Bolívar habría castigado semejante crimen cometido casi e	7.
	DCc_JRS_PLF3.txt	
55	Ahora los quiteños tratan de alzar una estatua a Bolívar . Dios les valga por generosos! La Biblia aconseja	7.
	DCc_JRS_PLF3.txt	
56	más no de la ajena, menos la de los ascendientes. Bolívar puso como chupa de dómine a los quiteños, en cart	7.
	DCc_JRS_PLF3.txt	
57	e los notables abogados don Elias Chaves y doctor Bolívar Chaves; y el doctor Lucas Vergara padre de tres b	8.
	DCc_JRS_PLF4.txt	
1	de Bolívar que el inca". Que en Bolívar brilló la justicia el mismo Unamuno, español, lo va a decir: "En mi	2.
	DCr_SEO_AC.txt	
2	pios y a extraños pero jamás con ensañamiento. Su justicia , como la de don Quijote, era rápida y ejecutiva.	2.
	DCr_SEO_AC.txt	
3	iéndole perder el castillo de Puerto Cabello" "La Justicia sola es la que conserva la República", decía (Car	2.
	DCr_SEO_AC.txt	
4	de Colombia, considerándolo—contra toda verdad y justicia — un terrible bellaco y delincuente afirman ser ju	3.
	DCr_SEO_IP2.txt	
5	hocaría tan directamente con el sentimiento de la justicia , que la conciencia humana no querría siquiera exa	3.
	DCr_SEO_IP2.txt	
6	tor". En primer lugar, porque no es amoldado a la justicia hacer responsables a los hijos de los delitos y c	3.
	DCr_SEO_IP2.txt	
7	to, sin decirlo, es responsable de sus actos. Con justicia esa refutación no fue publicada en Nariño. Hoy, a	7.
	DCc_JRS_PLF3.txt	

Ilustración 3 Contexto de "Bolívar" y "justicia" (fragmento)

Según este estudio del contexto de la palabra “justicia”, en algunos casos hace referencia al atributo del soberano (Bolívar elevado al rango de tal) y en otros se alude al sentimiento de rechazo frente a aquello que se considera atenta al respeto de lo instituido, o a la venganza por la profanación. Así en todos los casos.

B.4. Selección de las formas a conservar

Depurado el Glosario, se reduce, según el objetivo del estudio, quitando las palabras herramientas que no tienen significado en el Dominio del Corpus. Se obtiene así el glosario de formas (palabras y segmentos) dejando un listado

de 4826 formas (Tabla 4.). Se eliminaron artículos, conjunciones, preposiciones y pronombres y se dejó las palabras claves seleccionadas del dominio.

Tabla 4 Glosario de formas sin lematización (fragmento de 4826)

No	OCURRENCIAS	FORMAS
1	104	sañado
2	68	señor
3	65	doctor
4	57	bolívar
5	52	historia
6	40	obra
7	37	estudios
8	34	vida
9	32	ciudad
10	30	pasto
11	29	general
12	27	expiación
13	26	página
14	25	don
15	24	cabildo
16	24	nombre
17	23	modo
18	22	juan
19	22	sociedad
20	21	colombia
21	21	dios
22	21	nariño
23	20	hombres
24	20	ortiz
25	20	padre
26	20	verdad
27	19	gobernador
28	19	guerra
29	19	moral
30	17	años

B.2. Lematización y Desambiguación

Reduciendo el Glosario de formas (Tabla 4), se obtiene el Listado de Lemas.

Tabla 5 Ejemplo de lematización (fragmento)

á (132)
abanderado (1508) abanderado (*NCMS000) (0) abanderar (*VMPP0SM) (1)
abandonarlas (1509) abandonar (*VMN0000 [las]) (0)
abel (1510)
abiertamente (1511) abiertamente (*RG000) (0)
abnegado (1512) abnegado (*AGMS000) (0)
abofeteado (1513) abofetear (*VMPP0SM) (0)
abogado (781) abogado (*NCMS000) (0) abogar (*VMPP0SM) (1)
abogados (1514) abogado (*NCMP000 [os]) (0)
abominable (782) abominable (*AGIS000) (0)
aborígenes (1516) aborigen (*AGIP000) (0) aborigen (*NCIP000) (1)
aborrecimiento (1515)
abril (250) abril (*NCMS000) (0) abril (*NP00000) (1)
absoluta (1517) absolutamente (*RG000) (0) absoluto (*AGFS000) (1)
absolutamente (783) absolutamente (*RG000) (0)
absoluto (784) absoluto (*AGMS000) (0)
abstracción (1518) abstracción (*NCFS000) (0)
absurda (1519) absurdo (*AGFS000) (0)
abuelo (310) abuelo (*NCMS000) (0)
abuelos (311) abuelo (*NCMP000) (0)
abulia (1520)
abundan (1521) abundar (*VMIP3P0) (0)
abundante (785) abundante (*AGIS000) (0)
ac (1522)
acabamos (1523) acabar (*VMIS1P0) (0) acabar (*VMIP1P0) (1)
acabar (312) acabar (*VMN0000) (0)
acabe (1524) acabar (*VMR03S0) (0) acabar (*VMSP3S0) (1)
acabó (1525) acabar (*VMIS3S0) (0)
acabóse (1526) acabóse (*NCMS000) (0)
acaso (1527) acaso (*RG000) (0)

B.3. Análisis de la estructura léxica del vocabulario

A partir de este listado de lemas se realiza el estudio de la gama de frecuencias (Tabla 6) De este estudio, se elige el umbral de frecuencia (veinte 20), por cercanía al punto de inflexión de la curva de Gama de frecuencias, que podemos visualizar en la ilustración 4

Tabla 6. Gama de frecuencia de lemas

concepto	frecuencia	frecuencia relativa	porcentaje relativo
Saudo	104	1.0	0.010853212980872556
doctor	67	0.6633663366336634	0.007199656135826349
seor	66	0.6534653465346535	0.0070921985815602835
Bolvar	56	0.5544554455445545	0.006017623038899635
hombre	36	0.3564356435643564	0.0038684719535783366
Estudios	34	0.33663366336633666	0.0036535568450462067
historia	34	0.33663366336633666	0.0036535568450462067
vida	33	0.32673267326732675	0.0035460992907801418
Pasto	30	0.297029702970297	0.003223726627981947
palabra	27	0.26732673267326734	0.0029013539651837525
error	26	0.25742574257425743	0.0027938964109176876
nombre	26	0.25742574257425743	0.0027938964109176876
verdad	26	0.25742574257425743	0.0027938964109176876
pagina	26	0.25742574257425743	0.0027938964109176876
expiacion	25	0.24752475247524752	0.0026864388566516227
ciudad	25	0.24752475247524752	0.0026864388566516227
pueblo	24	0.2376237623762376	0.0025789813023855577
Total general	4244		

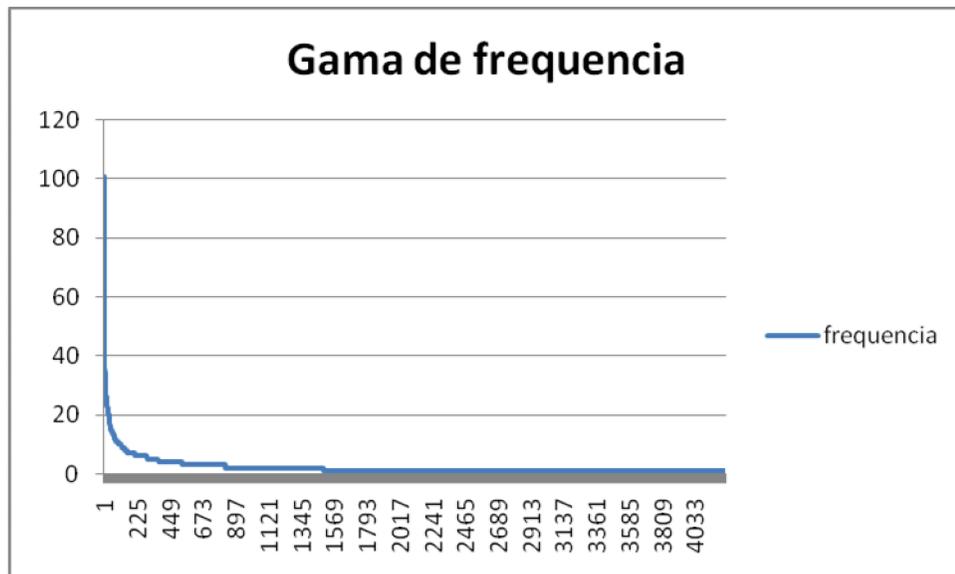


Ilustración 4 Gama de frecuencia de lemas

B.4. Selección de los lemas a conservar

Se seleccionan, entonces, los lemas que ocurrieron desde 20 veces en el corpus. Por ello la depuración del listado de lemas correspondería a 33 lemas. Se tomó la determinación de trabajar con los primeros 100 lemas, lo cual genera un amplio margen de inclusión. Se recuperaron frecuencias por errores de reconocimiento de texto plano (.txt)

Este glosario de lemas (Tabla 7), se utiliza en el análisis de la asociación de palabras. Para la creación de tablas léxicas es conveniente emplear el glosario de formas claves sin lematización, para no perder el vocabulario contexto de las mismas.

Tabla 7. Glosario de lemas seleccionados (100)

	FRECUENCIA	% MOSTRADO	% PROCESADO	% TOTAL	NO. CASOS	% CASOS	TF • IDF
<i>SAÑUDO</i>	104	5,00%	1,00%	0,50%	8	100,00%	0
<i>SEÑOR</i>	68	3,20%	0,70%	0,30%	5	62,50%	13,9
<i>DOCTOR</i>	65	3,10%	0,60%	0,30%	8	100,00%	0
<i>BOLÍVAR</i>	57	2,70%	0,50%	0,30%	7	87,50%	3,3
<i>HISTORIA</i>	52	2,50%	0,50%	0,20%	8	100,00%	0
<i>ERA</i>	44	2,10%	0,40%	0,20%	8	100,00%	0
<i>PUES</i>	43	2,10%	0,40%	0,20%	7	87,50%	2,5
<i>NI</i>	41	2,00%	0,40%	0,20%	8	100,00%	0
<i>OBRA</i>	40	1,90%	0,40%	0,20%	7	87,50%	2,3
<i>SINO</i>	40	1,90%	0,40%	0,20%	8	100,00%	0
<i>DICE</i>	38	1,80%	0,40%	0,20%	8	100,00%	0
<i>ESTUDIOS</i>	37	1,80%	0,40%	0,20%	8	100,00%	0
<i>LES</i>	34	1,60%	0,30%	0,20%	6	75,00%	4,2
<i>VIDA</i>	34	1,60%	0,30%	0,20%	7	87,50%	2
<i>CIUDAD</i>	32	1,50%	0,30%	0,10%	5	62,50%	6,5
<i>DECIR</i>	31	1,50%	0,30%	0,10%	8	100,00%	0
<i>PASTO</i>	30	1,40%	0,30%	0,10%	5	62,50%	6,1
<i>EXPIACIÓN</i>	27	1,30%	0,30%	0,10%	6	75,00%	3,4
<i>PÁGINA</i>	26	1,20%	0,30%	0,10%	4	50,00%	7,8
<i>DON</i>	25	1,20%	0,20%	0,10%	5	62,50%	5,1
<i>TAN</i>	25	1,20%	0,20%	0,10%	7	87,50%	1,4
<i>CABILDO</i>	24	1,10%	0,20%	0,10%	4	50,00%	7,2
<i>NOMBRE</i>	24	1,10%	0,20%	0,10%	7	87,50%	1,4
<i>MODO</i>	23	1,10%	0,20%	0,10%	7	87,50%	1,3
<i>JUAN</i>	22	1,00%	0,20%	0,10%	4	50,00%	6,6
<i>SOCIEDAD</i>	22	1,00%	0,20%	0,10%	4	50,00%	6,6
<i>COLOMBIA</i>	21	1,00%	0,20%	0,10%	7	87,50%	1,2
<i>DIOS</i>	21	1,00%	0,20%	0,10%	8	100,00%	0
<i>HECHO</i>	21	1,00%	0,20%	0,10%	6	75,00%	2,6
<i>NARIÑO</i>	21	1,00%	0,20%	0,10%	6	75,00%	2,6

HOMBRES	20	1,00%	0,20%	0,10%	7	87,50%	1,2
ORTIZ	20	1,00%	0,20%	0,10%	4	50,00%	6
PADRE	20	1,00%	0,20%	0,10%	7	87,50%	1,2
SIENDO	20	1,00%	0,20%	0,10%	8	100,00%	0
VERDAD	20	1,00%	0,20%	0,10%	6	75,00%	2,5
GOBERNADOR	19	0,90%	0,20%	0,10%	3	37,50%	8,1
GUERRA	19	0,90%	0,20%	0,10%	5	62,50%	3,9
HIZO	19	0,90%	0,20%	0,10%	6	75,00%	2,4
MORAL	19	0,90%	0,20%	0,10%	4	50,00%	5,7
DEBE	18	0,90%	0,20%	0,10%	7	87,50%	1
AÑOS	17	0,80%	0,20%	0,10%	6	75,00%	2,1
CIERTO	17	0,80%	0,20%	0,10%	7	87,50%	1
HIJOS	17	0,80%	0,20%	0,10%	5	62,50%	3,5
HOMBRE	17	0,80%	0,20%	0,10%	5	62,50%	3,5
SOLIDARIDAD	17	0,80%	0,20%	0,10%	1	12,50%	15,4
APUNTES	16	0,80%	0,20%	0,10%	4	50,00%	4,8
CAPITÁN	16	0,80%	0,20%	0,10%	2	25,00%	9,6
ENTONCES	16	0,80%	0,20%	0,10%	6	75,00%	2
MUCHOS	16	0,80%	0,20%	0,10%	8	100,00%	0
ALGUNOS	15	0,70%	0,10%	0,10%	6	75,00%	1,9
ERRORES	15	0,70%	0,10%	0,10%	6	75,00%	1,9
FRANCISCO	15	0,70%	0,10%	0,10%	4	50,00%	4,5
HABER	15	0,70%	0,10%	0,10%	6	75,00%	1,9
PALABRAS	15	0,70%	0,10%	0,10%	5	62,50%	3,1
TENÍA	15	0,70%	0,10%	0,10%	6	75,00%	1,9
DERECHO	14	0,70%	0,10%	0,10%	5	62,50%	2,9
ET	14	0,70%	0,10%	0,10%	3	37,50%	6
HACER	14	0,70%	0,10%	0,10%	7	87,50%	0,8
HUMANIDAD	14	0,70%	0,10%	0,10%	3	37,50%	6
LUGAR	14	0,70%	0,10%	0,10%	4	50,00%	4,2
MISMA	14	0,70%	0,10%	0,10%	6	75,00%	1,7
MUERTE	14	0,70%	0,10%	0,10%	6	75,00%	1,7
PATRIA	14	0,70%	0,10%	0,10%	5	62,50%	2,9
SANTANDER	14	0,70%	0,10%	0,10%	5	62,50%	2,9
SIGUIENTE	14	0,70%	0,10%	0,10%	5	62,50%	2,9
SIQUIERA	14	0,70%	0,10%	0,10%	5	62,50%	2,9
TIENEN	14	0,70%	0,10%	0,10%	7	87,50%	0,8
CADA	13	0,60%	0,10%	0,10%	6	75,00%	1,6
COSA	13	0,60%	0,10%	0,10%	5	62,50%	2,7
DIJO	13	0,60%	0,10%	0,10%	6	75,00%	1,6
DIO	13	0,60%	0,10%	0,10%	5	62,50%	2,7
HISTORIADOR	13	0,60%	0,10%	0,10%	6	75,00%	1,6
INDIO	13	0,60%	0,10%	0,10%	4	50,00%	3,9
NÚMERO	13	0,60%	0,10%	0,10%	4	50,00%	3,9
PECADOS	13	0,60%	0,10%	0,10%	2	25,00%	7,8
POPAYÁN	13	0,60%	0,10%	0,10%	3	37,50%	5,5
PRINCIPIO	13	0,60%	0,10%	0,10%	4	50,00%	3,9
SAN	13	0,60%	0,10%	0,10%	6	75,00%	1,6

<i>TODA</i>	13	0,60%	0,10%	0,10%	7	87,50%	0,8
<i>ESCRITO</i>	12	0,60%	0,10%	0,10%	6	75,00%	1,5
<i>HECHOS</i>	12	0,60%	0,10%	0,10%	7	87,50%	0,7
<i>HUBO</i>	12	0,60%	0,10%	0,10%	6	75,00%	1,5
<i>HUMANA</i>	12	0,60%	0,10%	0,10%	5	62,50%	2,4
<i>INDEPENDENCIA</i>	12	0,60%	0,10%	0,10%	6	75,00%	1,5
<i>JOSÉ</i>	12	0,60%	0,10%	0,10%	3	37,50%	5,1
<i>LIBERTADOR</i>	12	0,60%	0,10%	0,10%	4	50,00%	3,6
<i>PALABRA</i>	12	0,60%	0,10%	0,10%	4	50,00%	3,6
<i>PESOS</i>	12	0,60%	0,10%	0,10%	2	25,00%	7,2
<i>PUEBLO</i>	12	0,60%	0,10%	0,10%	6	75,00%	1,5
<i>PUEBLOS</i>	12	0,60%	0,10%	0,10%	4	50,00%	3,6
<i>PUEDEN</i>	12	0,60%	0,10%	0,10%	4	50,00%	3,6
<i>RESPONSABILIDAD</i>	12	0,60%	0,10%	0,10%	1	12,50%	10,8
<i>SEA</i>	12	0,60%	0,10%	0,10%	7	87,50%	0,7
<i>SOLDADOS</i>	12	0,60%	0,10%	0,10%	5	62,50%	2,4
<i>VARIOS</i>	12	0,60%	0,10%	0,10%	5	62,50%	2,4
<i>BOLETÍN</i>	11	0,50%	0,10%	0,10%	3	37,50%	4,7
<i>DEPARTAMENTO</i>	11	0,50%	0,10%	0,10%	4	50,00%	3,3
<i>JEFES</i>	11	0,50%	0,10%	0,10%	3	37,50%	4,7
<i>PARECE</i>	11	0,50%	0,10%	0,10%	4	50,00%	3,3
<i>REDENCIÓN</i>	11	0,50%	0,10%	0,10%	3	37,50%	4,7

C. Estudio de la tabla léxica base

La tabla léxica base recoge la distribución de formas entre los distintos textos individuales. Con la misma se confeccionan dos tablas léxicas base. En la primera, en filas se coloca los discursos y en columnas el vocabulario. La segunda es la “transpuesta”, es decir, se coloca en filas las formas claves seleccionadas (100) y en columnas los textos individuales correspondientes a cada discurso (8).

La primera tabla, se analiza por un análisis factorial de correspondencias simples (AFCS), para luego agrupar los discursos según su vocabulario y describirlos con datos complementarios. La segunda tabla, transpuesta, se utilizará más adelante para encontrar tipologías del vocabulario según su distribución en cada discurso.

C.1. Creación de la tabla léxica base

Análisis de la tabla léxica (discursos * vocabulario)

Se analizó la tabla léxica para agrupar los discursos según vocabulario y describir estos cluster en función del día y mes. En la siguiente ilustración 5, se observa un fragmento de la tabla léxica con las ocurrencias que representa la variable “léxica”.

FORMAS	Algunos	años	apuntes	boletín	Bolívar
ARTÍCULOS						
DCr_SEO_IP1	1	2			9	
DCr_SEO_AC	1	1			15	
DCr_SEO_IP2					6	
DCc_JRS_PG	1				7	
DCc_JRS_PLF1	5	3	3	1	10	
DCc_JRS_PLF2			7	3		
DCc_JRS_PLF3	4	4	2		9	
DCc_JRS_PLF4	3	4	4	7	1	

Ilustración 5 Fragmento de estructura léxica base (artículos*formas)

A esta tabla se le incluye, otras columnas que representan variables complementarias de tipo cualitativo o categórico correspondiente al carácter del artículo, autor, seudónimo. Las mismas no forman parte de la nube activa del análisis factorial, pero servirán para describir los cluster de discurso que se obtienen luego.

C.2. Representación gráfica

Esta tabla se analizó mediante análisis factorial [AFCS] tomando todas las formas y artículos como “activos”.

Este análisis, permitió descartar en una primera aproximación a la controversia entre Sergio Elías Ortiz y José Rafael Sañudo, la posibilidad de aplicar el análisis comparativo de textos a la réplicas y a los “Estudios sobre la vida de Bolívar”, porque ello reveló diferencias profundas en vocabulario y temáticas (Gráfico 4.2), lo cual influiría en el agrupamiento (cluster) de los mismos.

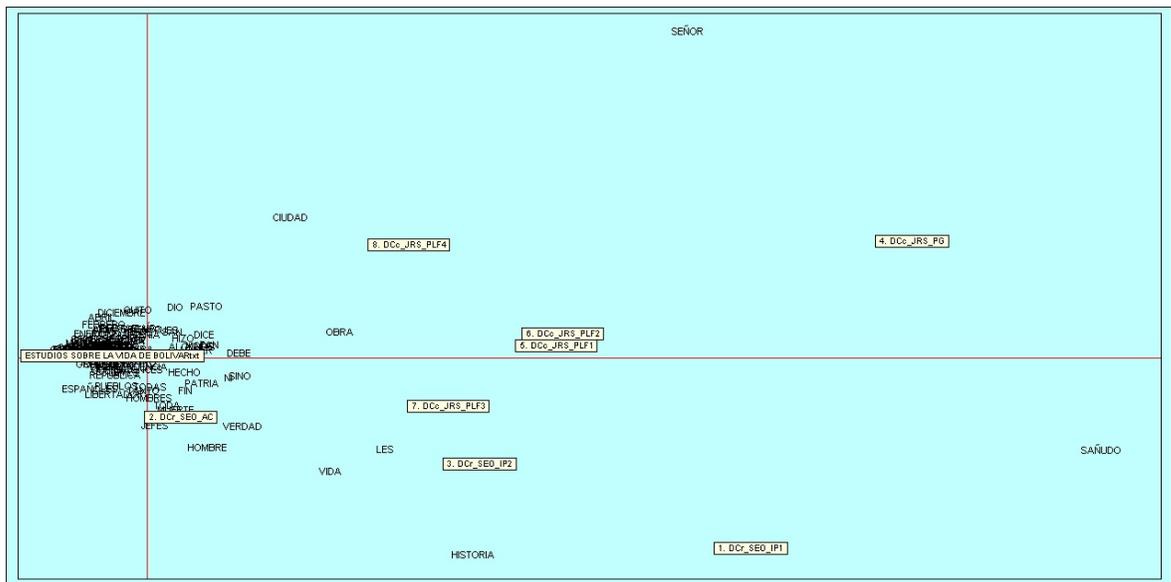


Ilustración 6 Discursos extremos "Estudios sobre la vida de Bolivar" y réplicas y contrarréplicas

Esto permitió detectar de manera más objetiva cuales son estos discursos “muy particulares”, se observó la inercia de cada discurso, se los ordenó de mayor a menos según la misma, (considerándose los de dispersión extrema respecto al resto, según podemos visualizar en la ilustración 6).

En la ilustración 7 se ha construido el árbol de agregación o Dendograma de clases a través del índice de similitud sobre los factores, podemos descubrir el aglomerado cluster de un solo elemento, que representa discursos con una marcada diferencia en su inercia.

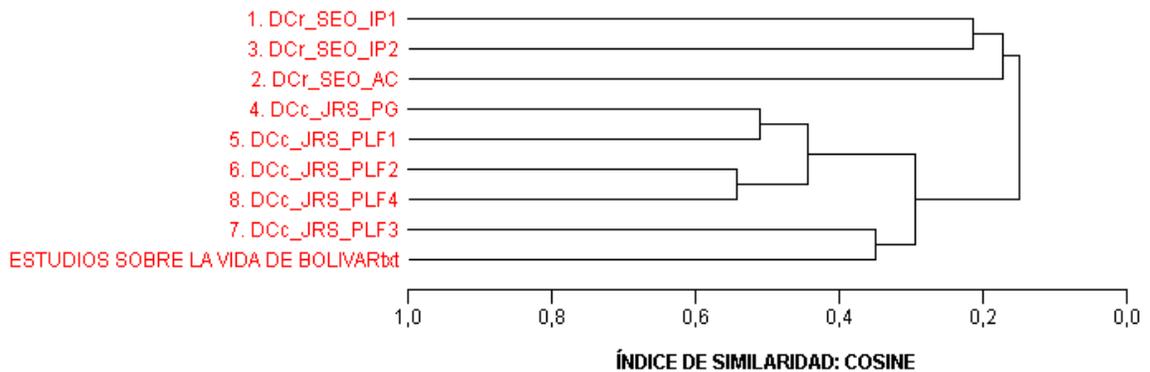


Ilustración 7 Distribución de la inercia de cada discurso

Se realizó nuevamente el AFCS, proyectando de manera ilustrativa los 8 artículos de la Revista “Ilustración Nariñense” correspondiente a las réplicas de Sergio Elías Ortiz frente a los Estudios sobre la vida de Bolívar y sus correspondientes contrarréplicas. En la ilustración 9., se puede observar la distribución de los mismos en el plano factorial de dos dimensiones⁶⁰.

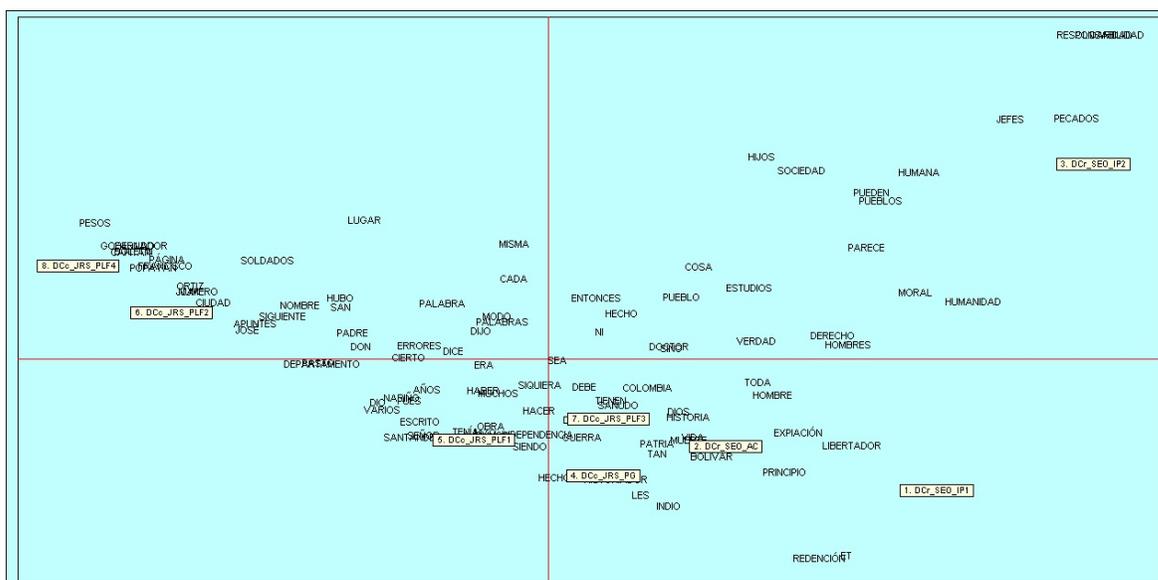


Ilustración 9 Distribución de los discursos según el vocabulario

Como era de esperar el análisis factorial de correspondencia simple reveló las particularidades en la utilización del vocabulario de cada uno de los artículos (tabla 9):

⁶⁰ Corresponde al plano formado por los factores 1 y 2 que recoge mayor información del hiperespacio.

Tabla 9. Análisis factorial de coocurrencia simple que muestra diferencias en la utilización de vocabulario en cada uno de los artículos (fragmento).

	1. DCr_S EO_IP	2. DCr _SE O_A C	3. DCr_S EO_IP	4. DCc_J RS_P	5. DCc_J RS_PL	6. DCc_JR S_PLF2	7. DCc_JR S_PLF3	8. DCc_J RS_PL	Chi2	P (2- colas)
	1	O_A C	2	G	F1			F4		
SAÑUDO	16	1	15	24	18	15	10	5	29,231	0
SEÑOR				32	12	5	3	16	103,53	0
DOCTOR	15	1	15	5	7	5	7	10	21,031	0,007
BOLÍVAR	9	15	6	7	10		9	1	23,421	0,003
HISTORIA	16	3	7	3	8	4	7	4	20	0,01
ERA	3	5	4	3	8	7	5	9	6,545	0,586
PUES	1	2		7	7	10	9	7	18,953	0,015
NI	4	3	9	3	7	4	4	7	6,805	0,558
OBRA	1		2	5	17	6	7	2	41,6	0
SINO	2	4	10	8	2	5	7	2	13,2	0,105
DICE	3	5	3	4	3	6	3	11	11,263	0,187
ESTUDIOS	3	1	14	2	3	3	9	2	30,676	0
LES	13			6	6	3	4	2	29,529	0
VIDA	6	4	4	3	5	2	10		14,471	0,07
CIUDAD				1	7	7	1	16	57	0
DECIR	2	1	3	7	6	3	7	2	10,548	0,229
PASTO				3	7	9	4	7	24,4	0,002
EXPIACIÓN	11		5	6		1	3	1	30,185	0
PÁGINA					1	5	3	17	73,692	0
DON	1	8		1		6		9	33,56	0
TAN	2	2	3	9	5	1	3		17,56	0,025
CABILDO					1	6	1	16	74	0
NOMBRE	1	1	1		2	11	2	6	32	0
MODO	1		4	2	3	4	4	5	7,261	0,509
JUAN				1	2	13		6	54,364	0
SOCIEDAD			13	2			3	4	50	0
COLOMBIA		2	4	5	3	1	5	1	9,857	0,275
DIOS	5	1	3	3	2	1	4	2	5,286	0,727
HECHO			6	1	6	3	4	1	16,714	0,033
NARIÑO		1		3	4	4	5	4	10,619	0,224

Al construir el árbol de agregación o Dendograma de clases (ilustración 10) se puede representar la agrupación de los artículos con su índice de similitud (calculado con el coeficiente de Jaccard, coocurrencia de palabra de primer orden).

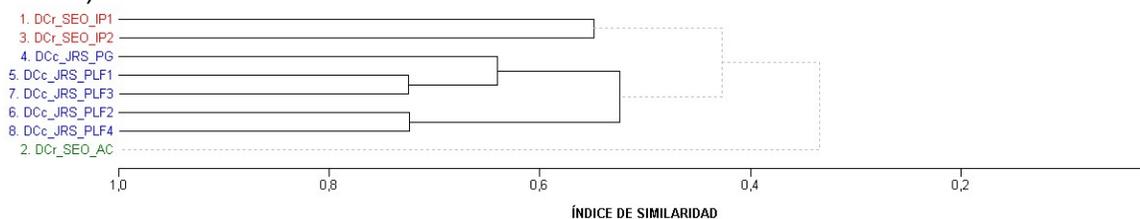


Ilustración 10 Dendograma de agrupación de los artículos estudiados

Se visualiza la coocurrencia de vocabulario en el gráfico de burbujas (ilustración 11), se calculó según el índice de similitud por el coeficiente de Jaccard (ocurrencia, tipo coocurrencia de palabra de primer orden)

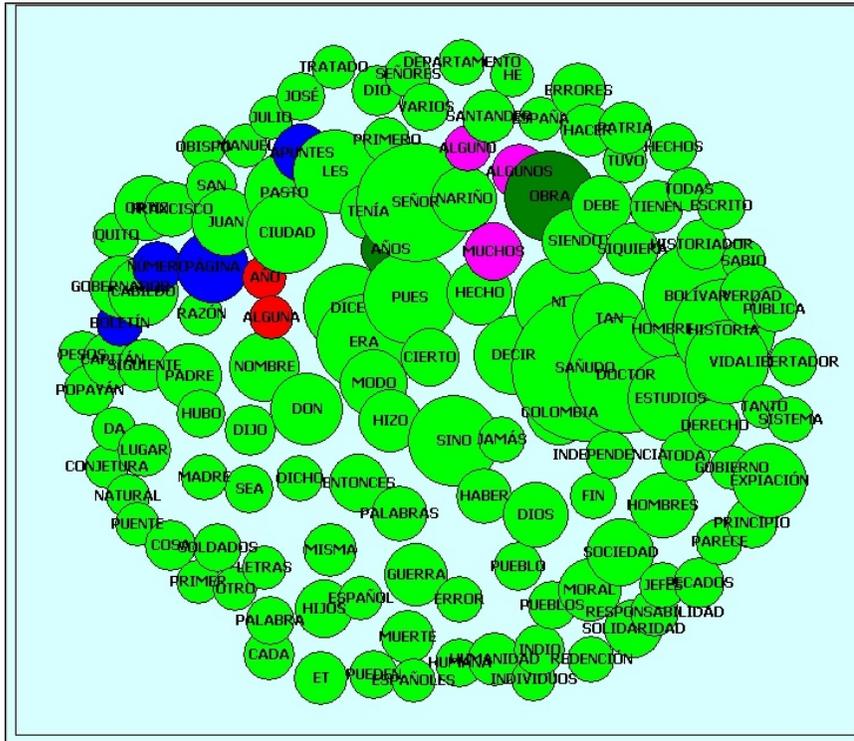


Ilustración 11 Gráfico de burbujas de cinco grupos por índice de similitud en vocabulario

En la ilustración 11 el tamaño de los puntos representa “relevancia” y está asociado a la contribución a los factores, los puntos más grandes son los que mayor información aportan al plano (este plano es el que captura la información más relevante de la tabla).

En la ilustración 12 se muestra el vocabulario y los artículos simultáneamente. Para facilitar la interpretación de este Mapa se selecciono solo el 20% de las formas de mayor relevancia para su visualización en la representación, según el criterio del Coseno cuadrado, como se ve en la ilustración 12. (Se visualiza el vocabulario según los ejes 2 y 3).

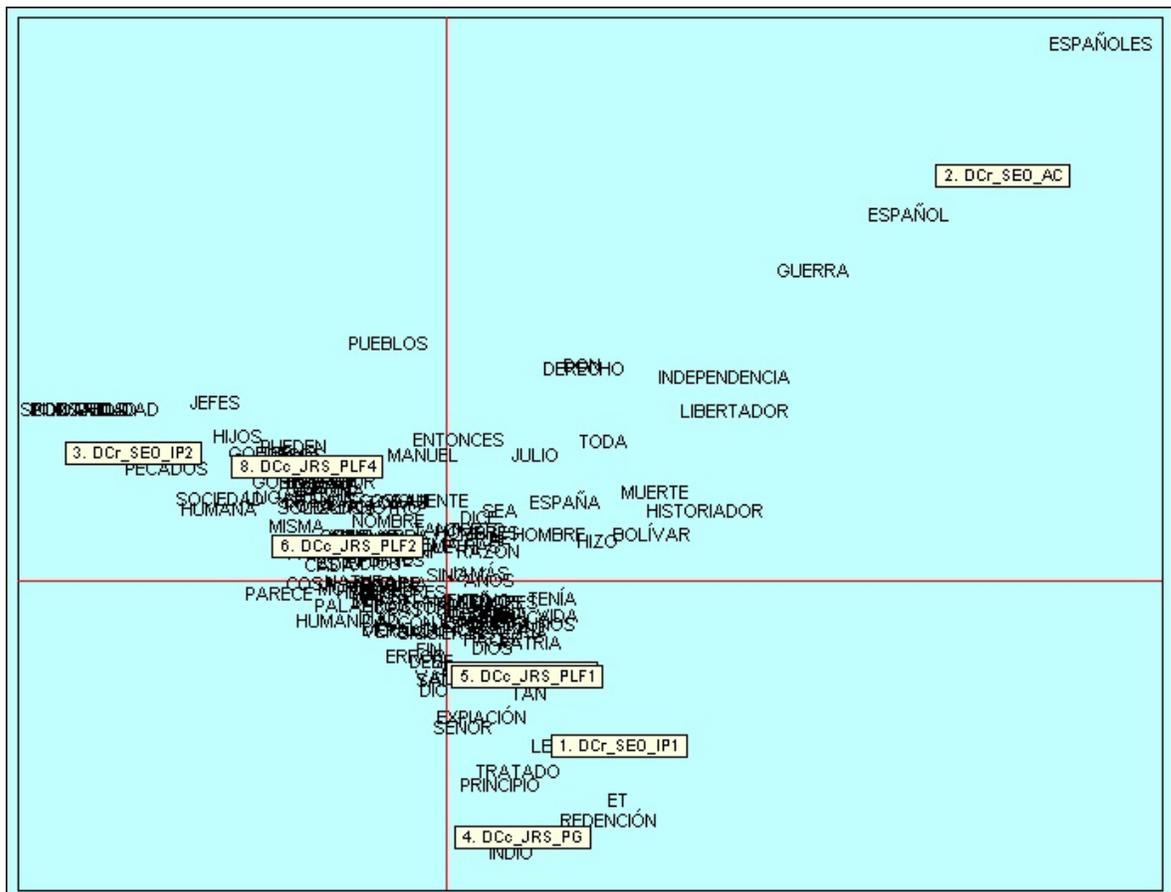


Ilustración 12 Mapa de artículos y vocabulario

Como ayuda a las interpretaciones se puede realizar un análisis de los ejes (complementario al AFCS) y a través del valor coeficiente de Jaccard, describir y asociar el vocabulario y artículos que caracterizan a cada eje que forma el plano principal.

Tabla 10 Análisis de correspondencia estadística para tabla cruzada por artículo

EIGENVALORES

Valores propios	Porcentajes	Acumulado
0,381	30,959	30,959
0,237	19,273	50,232
0,211	17,170	67,402

COORDENADAS DE VARIABLES

Elemento	Eje 1	Eje 2	Eje 3
----------	-------	-------	-------

1. DCr_SEO_IP1	0,679	0,393	-0,458
2. DCr_SEO_AC	0,298	1,264	1,09
3. DCr_SEO_IP2	1,043	-0,708	0,339
4. DCc_JRS_PG	0,057	0,176	-0,705
5. DCc_JRS_PLF1	-0,162	0,187	-0,267
6. DCc_JRS_PLF2	-0,626	-0,22	0,084
7. DCc_JRS_PLF3	0,056	0,175	-0,257
8. DCc_JRS_PLF4	-0,799	-0,311	0,3

PALABRAS/CATEGORIAS COORDINADAS

Elemento	Eje 1	Eje 2	Eje 3
SAÑUDO	0,314	-0,058	1,034
SEÑOR	-0,613	0,144	-1,485
DOCTOR	0,588	-0,276	-0,326
BOLÍVAR	0,706	1,676	0,386
HISTORIA	0,638	0,508	-0,565
ERA	-0,282	0,31	0,342
PUES	-0,661	0,261	-0,491
NI	0,267	-0,157	0,221
OBRA	-0,307	0,245	-1,012
SINO	0,577	0,006	-0,014
DICE	-0,414	0,269	0,561
ESTUDIOS	0,966	-0,714	0,101
LES	0,382	0,833	-1,667
VIDA	0,626	0,914	-0,402
CIUDAD	-1,494	-0,641	0,378
DECIR	0,096	0,303	-0,962
PASTO	-1,048	-0,228	-0,341
EXPIACIÓN	1,146	0,287	-1,397
PÁGINA	-1,689	-0,922	0,816
DON	-0,823	1,108	2,039
TAN	0,455	0,678	-1,153
CABILDO	-1,823	-1,044	0,943
NOMBRE	-1,081	-0,46	0,522
MODO	-0,205	-0,598	-0,104
JUAN	-1,576	-0,802	0,355
SOCIEDAD	1,272	-1,836	0,738
COLOMBIA	0,428	0,298	-0,382
DIOS	0,584	0,376	-0,717
HECHO	0,362	-0,647	-0,169

NARIÑO	-0,701	0,259	-0,416
HOMBRES	1,419	0,255	0,407
ORTIZ	-1,588	-0,823	0,368
PADRE	-0,872	-0,445	-0,097
SIENDO	-0,13	0,58	-0,955
VERDAD	0,948	-0,386	-0,549
GOBERNADOR	-1,814	-1,064	0,905
GUERRA	0,144	2,998	2,963
HIZO	-0,329	1,236	0,316
MORAL	1,747	-0,778	-0,136
DEBE	0,156	-0,113	-0,843
AÑOS	-0,569	0,35	-0,057
CIERTO	-0,623	0,24	0,446
HIJOS	1,124	-1,693	1,351
HOMBRE	1,024	0,848	0,393
SOLIDARIDAD	2,742	-2,987	1,606
APUNTES	-1,306	-0,494	0,139
CAPITÁN	-1,815	-1,073	0,781
ENTONCES	0,262	0,104	1,314
MUCHOS	-0,245	0,316	-0,275
ALGUNOS	-0,342	0,714	-0,486
ERRORES	-0,554	-0,321	-0,165
FRANCISCO	-1,687	-0,908	0,725
HABER	-0,311	0,248	-0,355
PALABRAS	-0,165	-0,663	-0,3
TENÍA	-0,421	0,867	-0,234
DERECHO	1,336	1,124	1,995
ET	1,317	1,398	-2,201
HACER	-0,057	0,395	-0,625
HUMANIDAD	2,03	-0,794	-0,449
LUGAR	-0,739	-1,388	0,755
MISMA	-0,061	-1,2	0,466
MUERTE	0,65	1,704	0,808
PATRIA	0,47	0,696	-0,673
SANTANDER	-0,67	0,359	-0,907
SIGUIENTE	-1,196	-0,166	0,711
SIQUIERA	-0,036	-0,06	-0,56
TIENEN	0,291	0,321	-0,462
CADA	-0,095	-0,938	0,095
COSA	0,753	-1,096	-0,086
DIJO	-0,291	-0,56	-0,419

DIO	-0,793	-0,085	-1,135
HISTORIADOR	0,242	2,105	0,626
INDIO	0,475	0,526	-2,706
NÚMERO	-1,555	-0,777	0,325
PECADOS	2,595	-2,272	1,025
POPAYÁN	-1,726	-0,945	0,663
PRINCIPIO	1,052	0,451	-2,055
SAN	-0,9	-0,301	0,713
TODA	0,976	1,282	1,287
ESCRITO	-0,616	0,515	-0,519
HECHOS	-0,006	0,979	-1,032
HUBO	-0,919	-0,7	0,205
HUMANA	1,827	-1,839	0,629
INDEPENDENCIA	-0,071	2,259	1,928
JOSÉ	-1,365	-0,405	0,217
LIBERTADOR	1,39	2,341	1,594
PALABRA	-0,412	-0,471	0,373
PESOS	-1,986	-1,218	1,165
PUEBLO	0,643	-0,457	0,337
PUEBLOS	1,632	-0,456	2,244
PUEDEN	1,612	-1,24	1,242
RESPONSABILIDAD	2,742	-2,987	1,606
SEA	0,057	0,448	0,617
SOLDADOS	-1,218	-0,979	0,651
VARIOS	-0,78	0,025	-1,002
ALGUNO	-0,345	-0,11	-0,541
BOLETÍN	-1,824	-1,018	0,898
DEPARTAMENTO	-1,028	-0,192	-0,257
DICHO	-0,555	-0,462	-0,527
ERROR	0,241	-0,236	-0,801
FIN	0,473	-0,133	-0,725
JAMÁS	-0,155	0,266	0,015
JEFES	2,276	-1,887	1,668
MADRE	-0,123	-0,517	-0,223
PARECE	1,55	-1,36	-0,196
PRIMERO	-0,477	0,482	-1,047
PÚBLICA	0,721	0,573	-0,413
QUITO	-1,619	-0,831	0,658
REDENCIÓN	1,19	1,326	-2,402
SEÑORES	-0,716	0,421	-0,265
SISTEMA	1,927	-0,197	0,295

ALGUNA	-0,55	-0,449	-0,097
AÑO	-1,225	-0,405	0,151
CONJETURA	-0,616	0,034	-0,47
DA	-0,596	0,456	-0,366
ESPAÑA	-0,473	0,974	0,699
ESPAÑOL	0,252	3,763	3,503
ESPAÑOLES	0,782	5,337	5,163
GOBIERNO	1,97	-1,404	1,18
HE	-0,929	0,441	0,35
INDIVIDUOS	2,742	-2,987	1,606
JULIO	-0,93	0,724	1,159
LETRAS	-1,024	0,17	0,27
MANUEL	-1,432	-0,19	1,156
NATURAL	0,776	-0,636	-0,074
OBISPO	-1,461	-0,567	0,308
OTRO	0,467	-0,399	0,645
PRIMER	-0,454	-1,045	0,206
PUENTE	-1,796	-0,988	0,845
RAZÓN	-0,715	0,347	0,223
SABIO	0,543	0,395	-0,387
TANTO	0,87	-0,027	0,431
TODAS	1,016	-0,39	0,704
TRATADO	-0,088	0,579	-1,91
TUVO	-0,181	0,337	-0,247

Según se observa en la Tabla 10, el eje 1, 2 y 3 dividen el espacio en torno a tres ejes, los cuales permiten visualizar la distribución en el plano del vocabulario (como palabras/categorías coordinadas) y los artículos (como coordenadas de variables).

Para la ayuda a la interpretación de estos gráficos, el estudio de especificidades de cada texto, el agrupamiento y descripción de cluster, son de gran ayuda para detalles puntuales y validaciones estadísticas del conocimiento inferido.

Obtención de tipologías

Se obtiene una matriz de factores de los discursos que representan la información contenida en esta tabla léxica, que guarda las relaciones entre discursos en función del vocabulario usado en ellos. Sobre esta matriz se aplica el procedimiento de clasificación jerárquica directa, método de Ward. En la siguiente ilustración 13, se visualiza el árbol o Dendograma con las diversas particiones realizadas.

Puede observarse en el gráfico Dendograma como cada discurso representa inicialmente una clase, y a través del proceso de agregación con el “vecino próximo”, se conforman los grupos homogéneos a distintos niveles, lo que permite obtener cluster por partición del árbol.

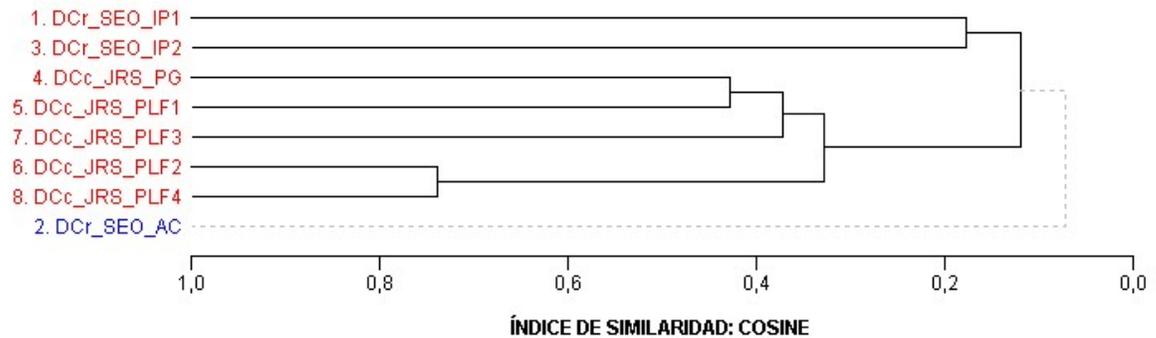


Ilustración 13 Dendograma - clasificación jerárquica directa de los artículos

Efectuado el corte del árbol en 2 grupos homogéneos (clase óptima según la relación de inercia intra dentro del grupo y la inercia inter entre clases. En la ilustración 14, están identificados los discursos agrupados y reclasificados en cada cluster.

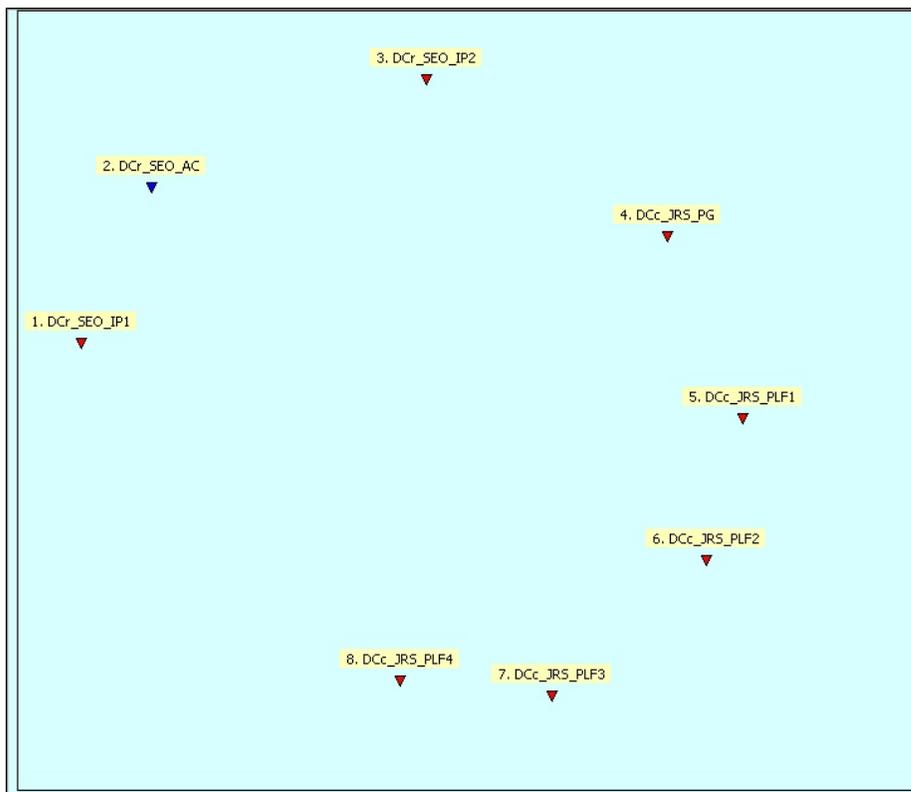


Ilustración 14 Discursos "activos" por grupo

En la ilustración 14, se identifican las diferentes clases, coloreando los grupos de discursos clasificados en cada grupo.

Si bien es cierto, que el proceso de clasificación por aglomeración, permite trabajar con grupos que manifiestan mayor inercia intra grupo, teniendo en cuenta que el análisis estadístico de los aglomerados de los artículos (tablas 11 y 12) señala que su índice de similaridad es bajo (en la matriz la pareja de artículos 8 y 6, de la tabla 11 y solo el nodo 3 en la tabla 12 tienen un índice de similaridad más cercano a la unidad), los artículos seleccionados por pueden mostrar mejor sus especificidades si se tienen en cuenta variables como carácter, autor, seudónimo.

Tabla 11 Matriz índice de similaridad

	1. DCr_SEO_IP1	2. DCr_SEO_AC	3. DCr_SEO_IP2	4. DCc_JRS_PG	5. DCc_JRS_PLF1	6. DCc_JRS_PLF2	7. DCc_JRS_PLF3	8. DCc_JRS_PLF4
1. DCr_SEO_IP1	1							
2. DCr_SEO_AC	0,08	1						
3. DCr_SEO_IP2	0,17	0,04	1					
4. DCc_JRS_PG	0,28	0,03	0,07	1				
5. DCc_JRS_PLF1	0,14	0,11	0,07	0,43	1			
6. DCc_JRS_PLF2	0,04	0,05	0,03	0,2	0,38	1		
7. DCc_JRS_PLF3	0,32	0,08	0,11	0,36	0,38	0,34	1	
8. DCc_JRS_PLF4	0,03	0,06	0,03	0,27	0,44	0,74	0,31	1

Tabla 12 Estadísticas aglomeración documentos

Nodo	Grupo 1	Grupo 2	Similaridad
1	4. DCc_JRS_PG	5. DCc_JRS_PLF1	0,426
2	Nodo 1	7. DCc_JRS_PLF3	0,368
3	6. DCc_JRS_PLF2	8. DCc_JRS_PLF4	0,742
4	Nodo 2	Nodo 3	0,324
5	1. DCr_SEO_IP1	3. DCr_SEO_IP2	0,169
6	Nodo 5	Nodo 4	0,11
7	Nodo 6	2. DCr_SEO_AC	0,063

En la tabla 13 se muestran las estadísticas de similaridad del vocabulario distribuido de acuerdo a los dos grupos aglomerados (cluster); en los dos grupos, el nodo 1 es el que presenta mayor similaridad:

Tabla 13 Estadísticas de similitud del vocabulario distribuido en los dos aglomerados de documentos

Nodo	Grupo 1	Grupo 2	Similaridad
1	RESPONSABILIDAD	SOLIDARIDAD	0,7
2	DOCTOR	SAÑUDO	0,534
3	CAPITÁN	PESOS	0,5
4	OBISPO	QUITO	0,5
5	NÚMERO	PÁGINA	0,471
6	FRANCISCO	GOBERNADOR	0,462
7	BOLETÍN	Nodo 5	0,433
8	JUAN	SAN	0,412
9	DON	GUERRA	0,4
10	CABILDO	PASTO	0,375
11	DEPARTAMENTO	RAZÓN	0,364
12	ENTONCES	HIZO	0,364
13	Nodo 8	MANUEL	0,359
14	INDIVIDUOS	SOCIEDAD	0,357
15	SIENDO	TUVO	0,353
16	NOMBRE	PUES	0,351
17	Nodo 10	CIUDAD	0,336
18	ALGUNA	AÑO	0,333
19	ALGUNOS	MUCHOS	0,333
20	LETRAS	PALABRA	0,333
21	ERA	NARIÑO	0,324
22	AÑOS	OBRA	0,321
23	MODO	TAN	0,321
24	DEBE	TODAS	0,316
25	DIJO	TENÍA	0,316
26	SIQUIERA	TIENEN	0,316
27	POPAYÁN	SIGUIENTE	0,313
28	SABIO	Nodo 24	0,313
29	CONJETURA	Nodo 28	0,308
30	JOSÉ	SANTANDER	0,308
31	MADRE	PRIMERO	0,308
32	BOLÍVAR	VIDA	0,317
33	Nodo 32	NI	0,246
34	Nodo 33	Nodo 2	0,222
35	Nodo 34	ESTUDIOS	0,207
36	Nodo 35	HISTORIA	0,185
37	Nodo 36	SINO	0,173
38	Nodo 37	SEÑOR	0,167
39	Nodo 38	DICE	0,175
40	Nodo 39	Nodo 26	0,171
41	Nodo 40	DECIR	0,173
42	Nodo 41	Nodo 16	0,174
43	Nodo 42	CIERTO	0,172
44	Nodo 43	Nodo 23	0,174
45	Nodo 44	Nodo 12	0,164
46	Nodo 45	COLOMBIA	0,158
47	Nodo 46	Nodo 25	0,157
48	Nodo 47	LES	0,145

49	Nodo 48	HOMBRE	0,142
50	Nodo 49	HECHO	0,14
51	Nodo 50	VERDAD	0,135
52	Nodo 51	Nodo 26	0,132
53	Nodo 52	Nodo 17	0,131
54	Nodo 53	DICHO	0,129
55	Nodo 54	JAMÁS	0,128
56	Nodo 55	HABER	0,127
57	Nodo 56	Nodo 15	0,124
58	Nodo 57	Nodo 24	0,123
59	Nodo 58	PALABRAS	0,12
60	Nodo 59	HACER	0,118
61	Nodo 60	HISTORIADOR	0,122
62	Nodo 61	Nodo 35	0,113
63	Nodo 62	Nodo 13	0,108
64	Nodo 63	Nodo 9	0,109
65	Nodo 64	Nodo 30	0,106
66	Nodo 65	Nodo 6	0,107
67	Nodo 66	PADRE	0,107
68	Nodo 67	Nodo 31	0,104
69	Nodo 68	HUBO	0,104
70	Nodo 69	VARIOS	0,104
71	Nodo 70	ESPAÑA	0,102
72	Nodo 71	Nodo 27	0,1
73	Nodo 72	ORTIZ	0,099
74	Nodo 73	Nodo 3	0,1
75	Nodo 74	LUGAR	0,101
76	Nodo 75	Nodo 4	0,097
77	Nodo 76	JULIO	0,092
78	Nodo 77	PATRIA	0,092
79	Nodo 78	SEA	0,092
80	Nodo 79	Nodo 62	0,09
81	Nodo 80	SEÑORES	0,09
82	Nodo 81	Nodo 11	0,088
83	Nodo 82	SOLDADOS	0,085
84	Nodo 83	MISMA	0,085
85	Nodo 84	Nodo 28	0,084
86	Nodo 85	DERECHO	0,086
87	Nodo 86	HOMBRES	0,082
88	Nodo 87	INDEPENDENCIA	0,082
89	Nodo 88	ESCRITO	0,082
90	Nodo 89	PUEBLO	0,08
91	Nodo 90	Nodo 50	0,078
92	Nodo 91	Nodo 29	0,077
93	Nodo 92	FIN	0,076
94	Nodo 93	HECHOS	0,075
95	Nodo 94	PÚBLICA	0,073
96	Nodo 95	ERRORES	0,07
97	Nodo 96	ESPAÑOL	0,068
98	Nodo 97	OTRO	0,069
99	Nodo 98	HIJOS	0,068
100	Nodo 99	Nodo 31	0,067

101	Nodo 100	NATURAL	0,066
102	Nodo 101	COSA	0,066
103	Nodo 102	GOBIERNO	0,065
104	Nodo 103	TRATADO	0,063
105	Nodo 104	Nodo 20	0,063
106	Nodo 105	Nodo 96	0,058
107	Nodo 106	TANTO	0,058
108	Nodo 107	MORAL	0,058
109	Nodo 108	PUEBLOS	0,056
110	Nodo 109	PUENTE	0,055
111	Nodo 110	LIBERTADOR	0,054
112	Nodo 111	MUERTE	0,054
113	Nodo 112	SISTEMA	0,048
114	Nodo 113	JEFES	0,047
115	Nodo 114	EXPIACIÓN	0,044
116	Nodo 115	PARECE	0,041
117	Nodo 116	Nodo 14	0,041
118	Nodo 117	HUMANIDAD	0,041
119	Nodo 118	HUMANA	0,042
120	Nodo 119	PECADOS	0,039
121	Nodo 120	Nodo 1	0,039
122	Nodo 121	PUEDEN	0,036
123	Nodo 122	PRINCIPIO	0,035
124	Nodo 123	INDIO	0,036
125	Nodo 124	ESPAÑOLES	0,033
126	Nodo 125	CADA	0,031
127	Nodo 126	REDENCIÓN	0,025
128	Nodo 127	Nodo 105	0,018
129	Nodo 19	Nodo 19	0,283
130	APUNTES	Nodo 7	0,262
131	Nodo 18	Nodo 130	0,21
132	Nodo 131	Nodo 22	0,192
133	Nodo 132	Nodo 129	0,122
134	Nodo 133	Nodo 128	0,099

C.3. Estudio de las especificidades

Dadas las razones señaladas, para este caso, es pertinente trabajar directamente con las variables identificadas (artículo, carácter, seudónimo, autor), las cuales generarían las clases buscadas.

DESCRIPCIÓN Y VALIDACIÓN ESTADÍSTICA DEL VOCABULARIO CARACTERÍSTICO DE CADA CLASE

Vocabulario característico por artículo

En la tabla 14 se presenta el análisis de vocabulario por documento organizado por frecuencia de palabras (se han incluido las 100 primeras formas sin lematizar, lo mismo que las que implican modales lógicos), presentado el número en cada columna por conteo de palabras que aparecen en el artículo:

Tabla 14 Tabla cruzada, presenta conteo de palabras por documento según el orden de su frecuencia de aparición

	1. DCr_ SEO_ IP1	2. DCr_ SEO_ AC	3. DCr_ SEO_ IP2	4. DCc_ JRS_ PG	5. DCc_ JRS_ PLF1	6. DCc_ JRS_ PLF2	7. DCc_ JRS_ PLF3	8. DCc_ JRS_ PLF4	Chi2	P (2- colas)
SAÑUDO	16	1	15	24	18	15	10	5	29,231	0
SEÑOR				32	12	5	3	16	103,53	0
DOCTOR	15	1	15	5	7	5	7	10	21,031	0,007
BOLÍVAR	9	15	6	7	10		9	1	23,421	0,003
HISTORIA	16	3	7	3	8	4	7	4	20	0,01
ERA	3	5	4	3	8	7	5	9	6,545	0,586
PUES	1	2		7	7	10	9	7	18,953	0,015
NI	4	3	9	3	7	4	4	7	6,805	0,558
OBRA	1		2	5	17	6	7	2	41,6	0
SINO	2	4	10	8	2	5	7	2	13,2	0,105
DICE	3	5	3	4	3	6	3	11	11,263	0,187
ESTUDIOS	3	1	14	2	3	3	9	2	30,676	0
LES	13			6	6	3	4	2	29,529	0
VIDA	6	4	4	3	5	2	10		14,471	0,07
CIUDAD				1	7	7	1	16	57	0
DECIR	2	1	3	7	6	3	7	2	10,548	0,229
PASTO				3	7	9	4	7	24,4	0,002
EXPIACIÓN	11		5	6		1	3	1	30,185	0
PÁGINA					1	5	3	17	73,692	0
DON	1	8		1		6		9	33,56	0
TAN	2	2	3	9	5	1	3		17,56	0,025
CABILDO					1	6	1	16	74	0
NOMBRE	1	1	1		2	11	2	6	32	0
MODO	1		4	2	3	4	4	5	7,261	0,509
JUAN				1	2	13		6	54,364	0
SOCIEDAD			13	2			3	4	50	0
COLOMBIA		2	4	5	3	1	5	1	9,857	0,275
DIOS	5	1	3	3	2	1	4	2	5,286	0,727
HECHO			6	1	6	3	4	1	16,714	0,033
NARIÑO		1		3	4	4	5	4	10,619	0,224
HOMBRES	5	3	7	2	1		1	1	16	0,042
ORTIZ				3	1	4		12	48	0
PADRE	1		1	2	3	4	2	7	13,6	0,093
SIENDO	1	1	1	4	8	2	2	1	16,8	0,032
VERDAD	2		6	2	1		8	1	24	0,002
GOBERNADOR						7	1	11	53	0
GUERRA	2	11				2	1	3	39,526	0
HIZO		4		3	3	3	4	2	7,526	0,481
MORAL	5		9		2		3		31,105	0
DEBE	1		3	3	6	2	2	1	10,444	0,235
AÑOS	2	1			3	3	4	4	8,882	0,352
CIERTO		2	1	1	3	6	2	2	10,765	0,215

HIJOS		1	10		1	4		1	39	0
HOMBRE	3	3	4		4		3		10,765	0,215
SOLIDARIDAD			17						119	0
APUNTES					3	7	2	4	23	0,003
CAPITÁN						10		6	52	0
ENTONCES		3	4		1	3	3	2	8	0,433
MUCHOS	1	1	1	1	3	4	4	1	7	0,537
ALGUNOS	1	1		1	5		4	3	13,267	0,103
ERRORES	2		1		2	6	2	2	13,267	0,103
FRANCISCO					1	5	1	8	33,533	0
HABER		1	1	2	2	5	4		12,2	0,143
PALABRAS			3	2	3	6	1		16,467	0,036
TENÍA		2		3	3	2	3	2	5,8	0,67
DERECHO	1	5	5		2		1		18	0,021
ET	10			2			2		47,714	0
HACER	1	1	1	3	1	4	3		7,714	0,462
HUMANIDAD	5		7	2					30,571	0
LUGAR			3	1		5		5	20,286	0,009
MISMA	1		4	1	1	3		4	11,143	0,194
MUERTE	5	4	1	1		1		2	13,429	0,098
PATRIA	2	1	2	1	8				28,286	0
SANTANDER				1	9	1	1	2	36,286	0
SIGUIENTE		1			1	3	3	6	18	0,021
SIQUIERA	1		2		7	3	1		22,571	0,004
TIENEN	2	1	2	2	3	3	1		4,286	0,83
CADA	1		3	2		2	1	4	8,538	0,383
COSA	1		5	3			1	3	14,692	0,065
DIJO			2	3	1	2	2	3	6,077	0,639
DIO				5	1	4	1	2	15,923	0,043
HISTORIADOR	1	4		1	2	1	4		11	0,202
INDIO	1		1	10			1		50,385	0
NÚMERO				2	1	2		8	31,923	0
PECADOS	2		11						63,923	0
POPAYÁN					1	7		5	33,154	0
PRINCIPIO	4		2	6			1		22,077	0,005
SAN		1	1		2	5	1	3	12,231	0,141
TODA	2	4	3	1	1		1	1	7,308	0,504
ESCRITO		1		3	3	2	1	2	6,667	0,573
HECHOS	2	1		3	3	1	1	1	5,333	0,721
HUBO			1	1	1	2	2	5	12	0,151
HUMANA	1		8	1			1	1	33,333	0
INDEPENDENCIA	1	5			2	1	1	2	12	0,151
JOSÉ					5	1		6	29,333	0
LIBERTADOR	4	5	2				1		18,667	0,017
PALABRA		1	2	2		7			26,667	0,001
PESOS						3		9	48	0
PUEBLO		1	4	2	1		2	2	8	0,433
PUEBLOS		3	7				1	1	28	0

PUEDEN	2	1	7					2	26,667	0,001
RESPONSABILIDAD			12						84	0
SEA		2	2	1	3	2	1	1	4	0,857
SOLDADOS			1		1	4	1	5	17,333	0,027
VARIOS				3	3	4	1	1	12	0,151
ALGUNO			1	1	3	1	3	2	7,182	0,517
BOLETÍN					1	3		7	31,909	0
DEPARTAMENTO					3	5	2	1	17,364	0,027
DICHO			1	3		2	2	3	8,636	0,374
ERROR			2	2	1	2	4		10,091	0,259
FIN	3		2	1	2	3			8,636	0,374
JAMÁS	1	1	1	1	1	3	2	1	2,818	0,945
JEFES		1	9		1				49,364	0
MADRE	1		2	1	3	2		2	5,727	0,678
PARECE	1		6	3		1			23,182	0,003
PRIMERO	1			1	7			2	29	0
PÚBLICA	1	1	2	1	2		4		8,636	0,374
QUITO					1	3	1	6	23,182	0,003
REDENCIÓN	7			3			1		31,909	0
SEÑORES		1		3	2		1	4	11,545	0,173
SISTEMA	4	1	5		1				20,273	0,009
ALGUNA			1	1	1		3	4	12,4	0,134
AÑO					2	2	2	4	12,4	0,134
CONJETURA						3	6	1	26,8	0,001
DA		1		3		2	2	2	7,6	0,473
ESPAÑA	1	2		1	1	1	1	3	4,4	0,819
ESPAÑOL		7			2			1	33,2	0
ESPAÑOLES		10							70	0
GOBIERNO		1	7	1	1				31,6	0
HE		1			5	1		3	18,8	0,016
INDIVIDUOS			10						70	0
JULIO		2			3	2		3	10,8	0,213
LETRAS		1		1	2	5		1	15,6	0,048
MANUEL		1			2	1		6	23,6	0,003
NATURAL	3		3		1	2		1	9,2	0,326
OBISPO					3	3		4	17,2	0,028
OTRO	1	1	3	1	1	1		2	4,4	0,819
PRIMER			2	1		5	1	1	15,6	0,048
PUENTE					1	3		6	26,8	0,001
RAZÓN	1	1		1	1	2	1	3	4,4	0,819
SABIO	1	1	2	2	3	1			6	0,647
TANTO	1	1	3		1		3	1	7,6	0,473
TODAS	1	1	4		3			1	12,4	0,134
TRATADO				4	1	1	4		17,2	0,028
TUVO		1	1	2	2	3	1		6	0,647

Se identificaron las características “relevantes” de cada documento, con validación estadística superior al 95% de certeza ($\alpha = 0.05$). Como quiera que en el cálculo del

chi-cuadrado obtuvimos un valor en cada fila, que supera al valor para $\alpha=0.005$, podremos concluir que las variables no son independientes, sino que están asociadas teniendo en cuenta que el valor del chi cuadrado es superior a p (un dato más extremo que el que proporciona el test o, equivalentemente, la probabilidad de obtener los datos observados si fuese cierta la hipótesis de independencia). Por lo tanto, a la vista de los resultados, rechazamos la hipótesis nula (H_0) y aceptamos la hipótesis alternativa (H_a) como probablemente cierta.

H_0 : No hay asociación entre las variables (en el ejemplo, los documentos entre sí por su vocabulario son independientes, no están asociados).

Y la hipótesis alternativa:

H_a : Sí hay asociación entre las variables, es decir, los documentos por su vocabulario están asociados.

Los valores de conteo de palabra ordenados por frecuencia en cada artículo, presentados en la tabla 14 se grafican a continuación (ilustración 15):

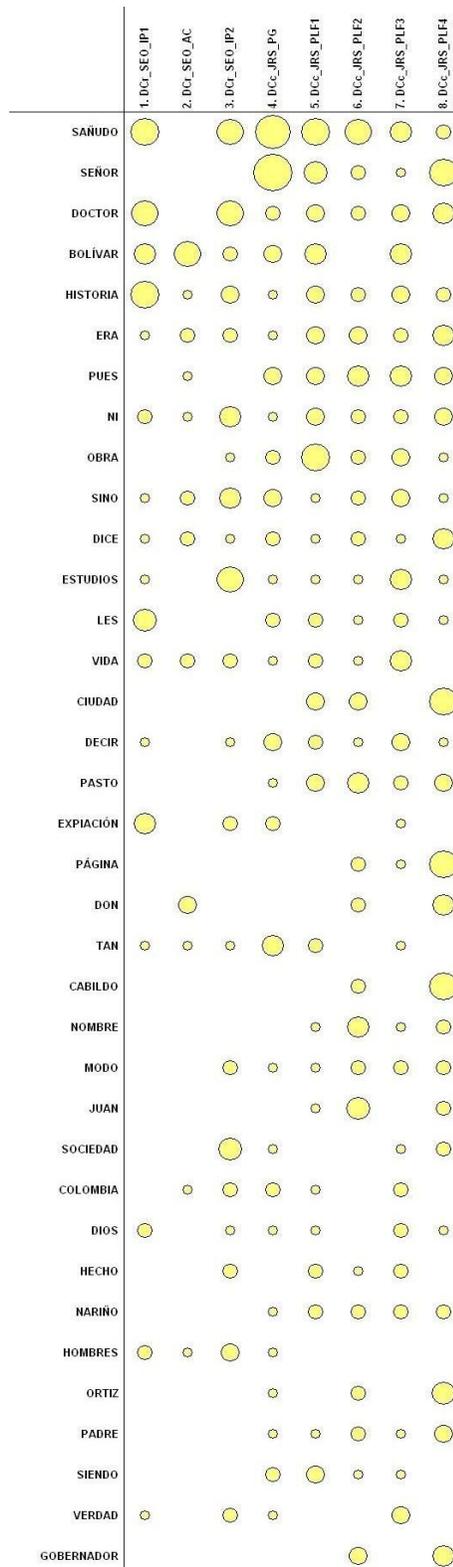


Ilustración 15 Gráfico de burbujas de palabras en cada documento (fragmento).

La Clasificación del vocabulario de los discursos permite obtener tipologías de los mismos en las que la progresión del vocabulario se produce en función del motivo y temáticas del artículo.

Palabras típicas y específicas de cada discurso.

Con base en la tabla 14 se elaboró la tabla 15 para establecer un listado ordenado con las unidades lexicales (palabras) típicas y exclusivas de cada artículo.

Las unidades lexicales "típicas" son definidas por exceso (sobre utilización) según un criterio estadístico que requiere la aplicación del Test de Chi cuadrado. Así puede observarse en la Tabla 16, las formas típicas definidas por valores altos del CHI2 y que conforman la "especificidad" del discurso (considerando valores de CHI2>20 en cada uno, se ven las formas más específicas).

Tabla 15 Tabla cruzada tabula frecuencia total con documentos presenta porcentaje de fila agrupadas por frecuencia de palabras.

	1. DCr_S EO_IP1	2. DCr_S EO_AC	3. DCr_S EO_IP2	4. DCc_J RS_PG	5. DCc_JRS _PLF1	6. DCc_JRS _PLF2	7. DCc_JRS _PLF3	8. DCc_JRS_P LF4	Chi2	P (2-colas)
SAÑUDO	15,40%	1,00%	14,40%	23,10%	17,30%	14,40%	9,60%	4,80%	29,231	0
SEÑOR				47,10%	17,60%	7,40%	4,40%	23,50%	103,53	0
DOCTOR	23,10%	1,50%	23,10%	7,70%	10,80%	7,70%	10,80%	15,40%	21,031	0,007
BOLÍVAR	15,80%	26,30%	10,50%	12,30%	17,50%		15,80%	1,80%	23,421	0,003
HISTORIA	30,80%	5,80%	13,50%	5,80%	15,40%	7,70%	13,50%	7,70%	20	0,01
ERA	6,80%	11,40%	9,10%	6,80%	18,20%	15,90%	11,40%	20,50%	6,545	0,586
PUES	2,30%	4,70%		16,30%	16,30%	23,30%	20,90%	16,30%	18,953	0,015
NI	9,80%	7,30%	22,00%	7,30%	17,10%	9,80%	9,80%	17,10%	6,805	0,558
OBRA	2,50%		5,00%	12,50%	42,50%	15,00%	17,50%	5,00%	41,6	0
SINO	5,00%	10,00%	25,00%	20,00%	5,00%	12,50%	17,50%	5,00%	13,2	0,105
DICE	7,90%	13,20%	7,90%	10,50%	7,90%	15,80%	7,90%	28,90%	11,263	0,187
ESTUDIOS	8,10%	2,70%	37,80%	5,40%	8,10%	8,10%	24,30%	5,40%	30,676	0
LES	38,20%			17,60%	17,60%	8,80%	11,80%	5,90%	29,529	0
VIDA	17,60%	11,80%	11,80%	8,80%	14,70%	5,90%	29,40%		14,471	0,07
CIUDAD				3,10%	21,90%	21,90%	3,10%	50,00%	57	0
DECIR	6,50%	3,20%	9,70%	22,60%	19,40%	9,70%	22,60%	6,50%	10,548	0,229
PASTO				10,00%	23,30%	30,00%	13,30%	23,30%	24,4	0,002
EXPIACIÓN	40,70%		18,50%	22,20%		3,70%	11,10%	3,70%	30,185	0
PÁGINA					3,80%	19,20%	11,50%	65,40%	73,692	0
DON	4,00%	32,00%		4,00%		24,00%		36,00%	33,56	0
TAN	8,00%	8,00%	12,00%	36,00%	20,00%	4,00%	12,00%		17,56	0,025
CABILDO					4,20%	25,00%	4,20%	66,70%	74	0
NOMBRE	4,20%	4,20%	4,20%		8,30%	45,80%	8,30%	25,00%	32	0
MODO	4,30%		17,40%	8,70%	13,00%	17,40%	17,40%	21,70%	7,261	0,509
JUAN				4,50%	9,10%	59,10%		27,30%	54,364	0
SOCIEDAD			59,10%	9,10%			13,60%	18,20%	50	0
COLOMBIA		9,50%	19,00%	23,80%	14,30%	4,80%	23,80%	4,80%	9,857	0,275

DIOS	23,80%	4,80%	14,30%	14,30%	9,50%	4,80%	19,00%	9,50%	5,286	0,727
HECHO			28,60%	4,80%	28,60%	14,30%	19,00%	4,80%	16,714	0,033
NARIÑO		4,80%		14,30%	19,00%	19,00%	23,80%	19,00%	10,619	0,224
HOMBRES	25,00%	15,00%	35,00%	10,00%	5,00%		5,00%	5,00%	16	0,042
ORTIZ				15,00%	5,00%	20,00%		60,00%	48	0
PADRE	5,00%		5,00%	10,00%	15,00%	20,00%	10,00%	35,00%	13,6	0,093
SIENDO	5,00%	5,00%	5,00%	20,00%	40,00%	10,00%	10,00%	5,00%	16,8	0,032
VERDAD	10,00%		30,00%	10,00%	5,00%		40,00%	5,00%	24	0,002
GOBERNADOR						36,80%	5,30%	57,90%	53	0
GUERRA	10,50%	57,90%				10,50%	5,30%	15,80%	39,526	0
HIZO		21,10%		15,80%	15,80%	15,80%	21,10%	10,50%	7,526	0,481
MORAL	26,30%		47,40%		10,50%		15,80%		31,105	0
DEBE	5,60%		16,70%	16,70%	33,30%	11,10%	11,10%	5,60%	10,444	0,235
AÑOS	11,80%	5,90%			17,60%	17,60%	23,50%	23,50%	8,882	0,352
CIERTO		11,80%	5,90%	5,90%	17,60%	35,30%	11,80%	11,80%	10,765	0,215
HIJOS		5,90%	58,80%		5,90%	23,50%		5,90%	39	0
HOMBRE	17,60%	17,60%	23,50%	100,00%	23,50%		17,60%		10,765	0,215
SOLIDARIDAD			%						119	0
APUNTES					18,80%	43,80%	12,50%	25,00%	23	0,003
CAPITÁN						62,50%		37,50%	52	0
ENTONCES		18,80%	25,00%		6,30%	18,80%	18,80%	12,50%	8	0,433
MUCHOS	6,30%	6,30%	6,30%	6,30%	18,80%	25,00%	25,00%	6,30%	7	0,537
ALGUNOS	6,70%	6,70%		6,70%	33,30%		26,70%	20,00%	13,267	0,103
ERRORES	13,30%		6,70%		13,30%	40,00%	13,30%	13,30%	13,267	0,103
FRANCISCO					6,70%	33,30%	6,70%	53,30%	33,533	0
HABER		6,70%	6,70%	13,30%	13,30%	33,30%	26,70%		12,2	0,143
PALABRAS			20,00%	13,30%	20,00%	40,00%	6,70%		16,467	0,036
TENÍA		13,30%		20,00%	20,00%	13,30%	20,00%	13,30%	5,8	0,67
DERECHO	7,10%	35,70%	35,70%		14,30%		7,10%		18	0,021
ET	71,40%			14,30%			14,30%		47,714	0
HACER	7,10%	7,10%	7,10%	21,40%	7,10%	28,60%	21,40%		7,714	0,462
HUMANIDAD	35,70%		50,00%	14,30%					30,571	0
LUGAR			21,40%	7,10%		35,70%		35,70%	20,286	0,009
MISMA	7,10%		28,60%	7,10%	7,10%	21,40%		28,60%	11,143	0,194
MUERTE	35,70%	28,60%	7,10%	7,10%		7,10%		14,30%	13,429	0,098
PATRIA	14,30%	7,10%	14,30%	7,10%	57,10%				28,286	0
SANTANDER				7,10%	64,30%	7,10%	7,10%	14,30%	36,286	0
SIGUIENTE		7,10%			7,10%	21,40%	21,40%	42,90%	18	0,021
SIQUIERA	7,10%		14,30%		50,00%	21,40%	7,10%		22,571	0,004
TIENEN	14,30%	7,10%	14,30%	14,30%	21,40%	21,40%	7,10%		4,286	0,83
CADA	7,70%		23,10%	15,40%		15,40%	7,70%	30,80%	8,539	0,383
COSA	7,70%		38,50%	23,10%			7,70%	23,10%	14,692	0,065
DIJO			15,40%	23,10%	7,70%	15,40%	15,40%	23,10%	6,077	0,639
DIO				38,50%	7,70%	30,80%	7,70%	15,40%	15,923	0,043
HISTORIADOR	7,70%	30,80%		7,70%	15,40%	7,70%	30,80%		11	0,202
INDIO	7,70%		7,70%	76,90%			7,70%		50,385	0
NÚMERO				15,40%	7,70%	15,40%		61,50%	31,923	0
PECADOS	15,40%		84,60%						63,923	0

POPAYÁN				7,70%	53,80%		38,50%	33,154	0	
PRINCIPIO	30,80%		15,40%	46,20%		7,70%		22,077	0,005	
SAN		7,70%	7,70%		15,40%	38,50%	7,70%	23,10%	12,231	0,141
TODA	15,40%	30,80%	23,10%	7,70%	7,70%		7,70%	7,70%	7,308	0,504
ESCRITO		8,30%		25,00%	25,00%	16,70%	8,30%	16,70%	6,667	0,573
HECHOS	16,70%	8,30%		25,00%	25,00%	8,30%	8,30%	8,30%	5,333	0,721
HUBO			8,30%	8,30%	8,30%	16,70%	16,70%	41,70%	12	0,151
HUMANA	8,30%		66,70%	8,30%			8,30%	8,30%	33,333	0
INDEPENDENCIA	8,30%	41,70%			16,70%	8,30%	8,30%	16,70%	12	0,151
JOSÉ					41,70%	8,30%		50,00%	29,333	0
LIBERTADOR	33,30%	41,70%	16,70%				8,30%		18,667	0,017
PALABRA		8,30%	16,70%	16,70%		58,30%			26,667	0,001
PESOS						25,00%		75,00%	48	0
PUEBLO		8,30%	33,30%	16,70%	8,30%		16,70%	16,70%	8	0,433
PUEBLOS		25,00%	58,30%				8,30%	8,30%	28	0
PUEDEN	16,70%	8,30%	58,30%	100,00%				16,70%	26,667	0,001
RESPONSABILIDAD									84	0
SEA		16,70%	16,70%	8,30%	25,00%	16,70%	8,30%	8,30%	4	0,857
SOLDADOS			8,30%		8,30%	33,30%	8,30%	41,70%	17,333	0,027
VARIOS				25,00%	25,00%	33,30%	8,30%	8,30%	12	0,151
ALGUNO			9,10%	9,10%	27,30%	9,10%	27,30%	18,20%	7,182	0,517
BOLETÍN					9,10%	27,30%		63,60%	31,909	0
DEPARTAMENTO					27,30%	45,50%	18,20%	9,10%	17,364	0,027
DICHO			9,10%	27,30%		18,20%	18,20%	27,30%	8,636	0,374
ERROR			18,20%	18,20%	9,10%	18,20%	36,40%		10,091	0,259
FIN	27,30%		18,20%	9,10%	18,20%	27,30%			8,636	0,374
JAMÁS	9,10%	9,10%	9,10%	9,10%	9,10%	27,30%	18,20%	9,10%	2,818	0,945
JEFES		9,10%	81,80%		9,10%				49,364	0
MADRE	9,10%		18,20%	9,10%	27,30%	18,20%		18,20%	5,727	0,678
PARECE	9,10%		54,50%	27,30%		9,10%			23,182	0,003
PRIMERO	9,10%			9,10%	63,60%			18,20%	29	0
PÚBLICA	9,10%	9,10%	18,20%	9,10%	18,20%		36,40%		8,636	0,374
QUITO					9,10%	27,30%	9,10%	54,50%	23,182	0,003
REDENCIÓN	63,60%			27,30%			9,10%		31,909	0
SEÑORES		9,10%		27,30%	18,20%		9,10%	36,40%	11,546	0,173
SISTEMA	36,40%	9,10%	45,50%		9,10%				20,273	0,009
ALGUNA			10,00%	10,00%	10,00%		30,00%	40,00%	12,4	0,134
AÑO					20,00%	20,00%	20,00%	40,00%	12,4	0,134
CONJETURA						30,00%	60,00%	10,00%	26,8	0,001
DA		10,00%		30,00%		20,00%	20,00%	20,00%	7,6	0,473
ESPAÑA	10,00%	20,00%		10,00%	10,00%	10,00%	10,00%	30,00%	4,4	0,819
ESPAÑOL		70,00%		20,00%				10,00%	33,2	0
ESPAÑOLES		100,00%							70	0
GOBIERNO		10,00%	70,00%	10,00%	10,00%				31,6	0
HE		10,00%			50,00%	10,00%		30,00%	18,8	0,016
INDIVIDUOS			100,00%						70	0
JULIO		20,00%			30,00%	20,00%		30,00%	10,8	0,213
LETRAS		10,00%		10,00%	20,00%	50,00%		10,00%	15,6	0,048

MANUEL					20,00%	10,00%		60,00%	23,6	0,003
NATURAL	30,00%		30,00%		10,00%	20,00%		10,00%	9,2	0,326
OBISPO					30,00%	30,00%		40,00%	17,2	0,028
OTRO	10,00%	10,00%	30,00%	10,00%	10,00%	10,00%		20,00%	4,4	0,819
PRIMER			20,00%	10,00%		50,00%	10,00%	10,00%	15,6	0,048
PUENTE					10,00%	30,00%		60,00%	26,8	0,001
RAZÓN	10,00%	10,00%		10,00%	10,00%	20,00%	10,00%	30,00%	4,4	0,819
SABIO	10,00%	10,00%	20,00%	20,00%	30,00%	10,00%			6	0,647
TANTO	10,00%	10,00%	30,00%		10,00%		30,00%	10,00%	7,6	0,473
TODAS	10,00%	10,00%	40,00%		30,00%			10,00%	12,4	0,134
TRATADO				40,00%	10,00%	10,00%	40,00%		17,2	0,028
TUVO		10,00%	10,00%	20,00%	20,00%	30,00%	10,00%		6	0,647

Tabla 16 Tabla cruzada tabula frecuencia total con documentos presenta porcentaje de fila agrupadas por frecuencia de palabras en orden descendente hasta un chi cuadrado 20

	1. DCr_S EO_IP1	2. DCr_S EO_AC	3. DCr_S EO_IP2 100,00 %	4. DCc_J RS_PG	5. DCc_JRS _PLF1	6. DCc_JRS _PLF2	7. DCc_JRS _PLF3	8. DCc_JRS_P LF4	Chi2	P (2-colas)
SOLIDARIDAD									119	0
SEÑOR				47,10%	17,60%	7,40%	4,40%	23,50%	103,53	0
RESPONSABILIDAD			100,00 %						84	0
CABILDO					4,20%	25,00%	4,20%	66,70%	74	0
PÁGINA					3,80%	19,20%	11,50%	65,40%	73,692	0
ESPAÑÓLES		100,00 %							70	0
INDIVIDUOS			100,00 %						70	0
PECADOS	15,40%		84,60%						63,923	0
CIUDAD				3,10%	21,90%	21,90%	3,10%	50,00%	57	0
JUAN				4,50%	9,10%	59,10%		27,30%	54,364	0
GOBERNADOR						36,80%	5,30%	57,90%	53	0
CAPITÁN						62,50%		37,50%	52	0
INDIO	7,70%		7,70%	76,90%			7,70%		50,385	0
SOCIEDAD			59,10%	9,10%			13,60%	18,20%	50	0
JEFES		9,10%	81,80%		9,10%				49,364	0
ORTIZ				15,00%	5,00%	20,00%		60,00%	48	0
PESOS						25,00%		75,00%	48	0
ET	71,40%			14,30%			14,30%		47,714	0
OBRA	2,50%		5,00%	12,50%	42,50%	15,00%	17,50%	5,00%	41,6	0
GUERRA	10,50%	57,90%				10,50%	5,30%	15,80%	39,526	0
HIJOS		5,90%	58,80%		5,90%	23,50%		5,90%	39	0
SANTANDER				7,10%	64,30%	7,10%	7,10%	14,30%	36,286	0
DON	4,00%	32,00%		4,00%		24,00%		36,00%	33,56	0
FRANCISCO					6,70%	33,30%	6,70%	53,30%	33,533	0
HUMANA	8,30%		66,70%	8,30%			8,30%	8,30%	33,333	0
ESPAÑÓL		70,00%			20,00%			10,00%	33,2	0
POPAYÁN					7,70%	53,80%		38,50%	33,154	0
NOMBRE	4,20%	4,20%	4,20%		8,30%	45,80%	8,30%	25,00%	32	0
NÚMERO				15,40%	7,70%	15,40%		61,50%	31,923	0

BOLETÍN				9,10%	27,30%		63,60%	31,909	0	
REDENCIÓN	63,60%		27,30%			9,10%		31,909	0	
GOBIERNO		10,00%	70,00%	10,00%	10,00%			31,6	0	
MORAL	26,30%		47,40%		10,50%		15,80%	31,105	0	
ESTUDIOS	8,10%	2,70%	37,80%	5,40%	8,10%	8,10%	24,30%	5,40%	30,676	0
HUMANIDAD	35,70%		50,00%	14,30%					30,571	0
EXPIACIÓN	40,70%		18,50%	22,20%		3,70%	11,10%	3,70%	30,185	0
LES	38,20%		17,60%	17,60%	8,80%	11,80%	5,90%	29,529	0	
JOSÉ				41,70%	8,30%		50,00%	29,333	0	
SAÑUDO	15,40%	1,00%	14,40%	23,10%	17,30%	14,40%	9,60%	4,80%	29,231	0
PRIMERO	9,10%			9,10%	63,60%			18,20%	29	0
PATRIA	14,30%	7,10%	14,30%	7,10%	57,10%				28,286	0
PUEBLOS		25,00%	58,30%				8,30%	8,30%	28	0
CONJETURA						30,00%	60,00%	10,00%	26,8	0,001
PUENTE				10,00%	30,00%			60,00%	26,8	0,001
PALABRA		8,30%	16,70%	16,70%		58,30%			26,667	0,001
PUEDEN	16,70%	8,30%	58,30%					16,70%	26,667	0,001
PASTO				10,00%	23,30%	30,00%	13,30%	23,30%	24,4	0,002
VERDAD	10,00%		30,00%	10,00%	5,00%		40,00%	5,00%	24	0,002
MANUEL		10,00%			20,00%	10,00%		60,00%	23,6	0,003
BOLÍVAR	15,80%	26,30%	10,50%	12,30%	17,50%		15,80%	1,80%	23,421	0,003
PARECE	9,10%		54,50%	27,30%		9,10%			23,182	0,003
QUITO				9,10%	27,30%	9,10%	54,50%		23,182	0,003
APUNTES				18,80%	43,80%	12,50%	25,00%		23	0,003
SIQUIERA	7,10%		14,30%		50,00%	21,40%	7,10%		22,571	0,004
PRINCIPIO	30,80%		15,40%	46,20%			7,70%		22,077	0,005
DOCTOR	23,10%	1,50%	23,10%	7,70%	10,80%	7,70%	10,80%	15,40%	21,031	0,007
LUGAR			21,40%	7,10%		35,70%		35,70%	20,286	0,009
SISTEMA	36,40%	9,10%	45,50%		9,10%				20,273	0,009
HISTORIA	30,80%	5,80%	13,50%	5,80%	15,40%	7,70%	13,50%	7,70%	20	0,01

Tabla 17 Presenta las palabras típicas de Sergio Elías Ortíz en el primer artículo “HISTORIA?” con el seudónimo “Un indio patriota”

	1. DCr_S EO_IP1	Chi2	P (2-colas)
ET	71,40%	47,714	0
REDENCIÓN	63,60%	31,909	0
EXPIACIÓN	40,70%	30,185	0
LES	38,20%	29,529	0
SISTEMA	36,40%	20,273	0,009
HUMANIDAD	35,70%	30,571	0
PRINCIPIO	30,80%	22,077	0,005
HISTORIA	30,80%	20	0,01
MORAL	26,30%	31,105	0
DOCTOR	23,10%	21,031	0,007
PUEDEN	16,70%	26,667	0,001
BOLÍVAR	15,80%	23,421	0,003

PECADOS	15,40%	63,923	0
SAÑUDO	15,40%	29,231	0
PATRIA	14,30%	28,286	0
GUERRA	10,50%	39,526	0
VERDAD	10,00%	24	0,002
PRIMERO	9,10%	29	0
PARECE	9,10%	23,182	0,003
HUMANA	8,30%	33,333	0
ESTUDIOS	8,10%	30,676	0
INDIO	7,70%	50,385	0
SIQUIERA	7,10%	22,571	0,004
NOMBRE	4,20%	32	0
DON	4,00%	33,56	0
OBRA	2,50%	41,6	0

Tabla 18. Presenta las palabras típicas de Sergio Elías Ortíz en el segundo artículo “AÚN VIVE LA HISTORIA” con el seudónimo “Alexo Corfú”

	2. DCr_SEO _AC	Chi2	P (2-colas)
ESPAÑOLES	100,00%	70	0
ESPAÑOL	70,00%	33,2	0
GUERRA	57,90%	39,526	0
DON	32,00%	33,56	0
BOLÍVAR	26,30%	23,421	0,003
PUEBLOS	25,00%	28	0
GOBIERNO	10,00%	31,6	0
MANUEL	10,00%	23,6	0,003
SISTEMA	9,10%	20,273	0,009
JEFES	9,10%	49,364	0
PUEDEN	8,30%	26,667	0,001
PALABRA	8,30%	26,667	0,001
PATRIA	7,10%	28,286	0
HIJOS	5,90%	39	0
HISTORIA	5,80%	20	0,01
NOMBRE	4,20%	32	0
ESTUDIOS	2,70%	30,676	0
DOCTOR	1,50%	21,031	0,007
SAÑUDO	1,00%	29,231	0

Tabla 19. Presenta las palabras típicas de Sergio Elías Ortíz en el tercer artículo “*HISTORIA?*” con el seudónimo “Un indio patriota”

	3. DCr_SEO _IP2	Chi2	P (2-colas)
SOLIDARIDAD	100,00%	119	0
RESPONSABILIDAD	100,00%	84	0
INDIVIDUOS	100,00%	70	0
PECADOS	84,60%	63,923	0
JEFES	81,80%	49,364	0
GOBIERNO	70,00%	31,6	0
HUMANA	66,70%	33,333	0
SOCIEDAD	59,10%	50	0
HIJOS	58,80%	39	0
PUEDEN	58,30%	26,667	0,001
PUEBLOS	58,30%	28	0
PARECE	54,50%	23,182	0,003
HUMANIDAD	50,00%	30,571	0
MORAL	47,40%	31,105	0
SISTEMA	45,50%	20,273	0,009
ESTUDIOS	37,80%	30,676	0
VERDAD	30,00%	24	0,002
DOCTOR	23,10%	21,031	0,007
LUGAR	21,40%	20,286	0,009
EXPIACIÓN	18,50%	30,185	0
PALABRA	16,70%	26,667	0,001
PRINCIPIO	15,40%	22,077	0,005
SAÑUDO	14,40%	29,231	0
PATRIA	14,30%	28,286	0
SIQUIERA	14,30%	22,571	0,004
HISTORIA	13,50%	20	0,01
BOLÍVAR	10,50%	23,421	0,003
INDIO	7,70%	50,385	0
OBRA	5,00%	41,6	0
NOMBRE	4,20%	32	0

Tabla 20. Presenta las palabras típicas de José Rafael Sañudo en el artículo “Los refutadores del señor José Rafael Sañudo” con el seudónimo “Próspero Gallo”

	4. DCc_J RS_PG	Chi2	P (2-colas)
INDIO	76,90%	50,385	0
SEÑOR	47,10%	103,53	0
PRINCIPIO	46,20%	22,077	0,005
REDENCIÓN	27,30%	31,909	0
PARECE	27,30%	23,182	0,003
SAÑUDO	23,10%	29,231	0
EXPIACIÓN	22,20%	30,185	0
LES	17,60%	29,529	0
PALABRA	16,70%	26,667	0,001
NÚMERO	15,40%	31,923	0
ORTIZ	15,00%	48	0
ET	14,30%	47,714	0
HUMANIDAD	14,30%	30,571	0
OBRA	12,50%	41,6	0
BOLÍVAR	12,30%	23,421	0,003
VERDAD	10,00%	24	0,002
GOBIERNO	10,00%	31,6	0
PASTO	10,00%	24,4	0,002
PRIMERO	9,10%	29	0
SOCIEDAD	9,10%	50	0
HUMANA	8,30%	33,333	0
DOCTOR	7,70%	21,031	0,007
PATRIA	7,10%	28,286	0
SANTANDER	7,10%	36,286	0
LUGAR	7,10%	20,286	0,009
HISTORIA	5,80%	20	0,01
ESTUDIOS	5,40%	30,676	0
JUAN	4,50%	54,364	0
DON	4,00%	33,56	0
CIUDAD	3,10%	57	0

Tabla 21. Presenta las palabras típicas de José Rafael Sañudo en el artículo “UNAS VERDADES (1)” con el seudónimo “Patricio Leal y Franco”

	5. DCc_JRS _PLF1	Chi2	P (2-colas)
SANTANDER	64,30%	36,286	0
PRIMERO	63,60%	29	0
PATRIA	57,10%	28,286	0
SIQUIERA	50,00%	22,571	0,004
OBRA	42,50%	41,6	0
JOSÉ	41,70%	29,333	0
PASTO	23,30%	24,4	0,002
CIUDAD	21,90%	57	0
ESPAÑOL	20,00%	33,2	0
MANUEL	20,00%	23,6	0,003
APUNTES	18,80%	23	0,003
LES	17,60%	29,529	0
SEÑOR	17,60%	103,53	0
BOLÍVAR	17,50%	23,421	0,003
SAÑUDO	17,30%	29,231	0
HISTORIA	15,40%	20	0,01
DOCTOR	10,80%	21,031	0,007
MORAL	10,50%	31,105	0
GOBIERNO	10,00%	31,6	0
PUENTE	10,00%	26,8	0,001
SISTEMA	9,10%	20,273	0,009
JUAN	9,10%	54,364	0
JEFES	9,10%	49,364	0
BOLETÍN	9,10%	31,909	0
QUITO	9,10%	23,182	0,003
NOMBRE	8,30%	32	0
ESTUDIOS	8,10%	30,676	0
POPAYÁN	7,70%	33,154	0
NÚMERO	7,70%	31,923	0
FRANCISCO	6,70%	33,533	0
HIJOS	5,90%	39	0
VERDAD	5,00%	24	0,002
ORTIZ	5,00%	48	0
CABILDO	4,20%	74	0
PÁGINA	3,80%	73,692	0

Tabla 22. Presenta las palabras típicas de José Rafael Sañudo en el artículo “UNAS VERDADES (2)” con el seudónimo “Patricio Leal y Franco”

	6. DCc_JRS _PLF2	Chi2	P (2-colas)
CAPITÁN	62,50%	52	0
JUAN	59,10%	54,364	0
PALABRA	58,30%	26,667	0,001
POPAYÁN	53,80%	33,154	0
NOMBRE	45,80%	32	0
APUNTES	43,80%	23	0,003
GOBERNADOR	36,80%	53	0
LUGAR	35,70%	20,286	0,009
FRANCISCO	33,30%	33,533	0
CONJETURA	30,00%	26,8	0,001
PUENTE	30,00%	26,8	0,001
PASTO	30,00%	24,4	0,002
BOLETÍN	27,30%	31,909	0
QUITO	27,30%	23,182	0,003
CABILDO	25,00%	74	0
PESOS	25,00%	48	0
DON	24,00%	33,56	0
HIJOS	23,50%	39	0
CIUDAD	21,90%	57	0
SIQUIERA	21,40%	22,571	0,004
ORTIZ	20,00%	48	0
PÁGINA	19,20%	73,692	0
NÚMERO	15,40%	31,923	0
OBRA	15,00%	41,6	0
SAÑUDO	14,40%	29,231	0
GUERRA	10,50%	39,526	0
MANUEL	10,00%	23,6	0,003
PARECE	9,10%	23,182	0,003
LES	8,80%	29,529	0
JOSÉ	8,30%	29,333	0
ESTUDIOS	8,10%	30,676	0
HISTORIA	7,70%	20	0,01
DOCTOR	7,70%	21,031	0,007
SEÑOR	7,40%	103,53	0
SANTANDER	7,10%	36,286	0
EXPIACIÓN	3,70%	30,185	0

Tabla 23. Presenta las palabras típicas de José Rafael Sañudo en el artículo “UNAS VERDADES (3)” con el seudónimo “Patricio Leal y Franco”

	7. DCC_JRS _PLF3	Chi2	P (2-colas)
CONJETURA	60,00%	26,8	0,001
VERDAD	40,00%	24	0,002
ESTUDIOS	24,30%	30,676	0
OBRA	17,50%	41,6	0
MORAL	15,80%	31,105	0
BOLÍVAR	15,80%	23,421	0,003
ET	14,30%	47,714	0
SOCIEDAD	13,60%	50	0
HISTORIA	13,50%	20	0,01
PASTO	13,30%	24,4	0,002
APUNTES	12,50%	23	0,003
LES	11,80%	29,529	0
PÁGINA	11,50%	73,692	0
EXPIACIÓN	11,10%	30,185	0
DOCTOR	10,80%	21,031	0,007
SAÑUDO	9,60%	29,231	0
REDENCIÓN	9,10%	31,909	0
QUITO	9,10%	23,182	0,003
HUMANA	8,30%	33,333	0
NOMBRE	8,30%	32	0
PUEBLOS	8,30%	28	0
PRINCIPIO	7,70%	22,077	0,005
INDIO	7,70%	50,385	0
SIQUIERA	7,10%	22,571	0,004
SANTANDER	7,10%	36,286	0
FRANCISCO	6,70%	33,533	0
GUERRA	5,30%	39,526	0
GOBERNADOR	5,30%	53	0
SEÑOR	4,40%	103,53	0
CABILDO	4,20%	74	0
CIUDAD	3,10%	57	0

Tabla 24. Presenta las palabras típicas de José Rafael Sañudo en el artículo “UNAS VERDADES (4)” con el seudónimo “Patricio Leal y Franco”

	8. DCc_JRS_P LF4	Chi2	P (2-colas)
PESOS	75,00%	48	0
CABILDO	66,70%	74	0
PÁGINA	65,40%	73,692	0
BOLETÍN	63,60%	31,909	0
NÚMERO	61,50%	31,923	0
ORTIZ	60,00%	48	0
PUENTE	60,00%	26,8	0,001
MANUEL	60,00%	23,6	0,003
GOBERNADOR	57,90%	53	0
QUITO	54,50%	23,182	0,003
FRANCISCO	53,30%	33,533	0
CIUDAD	50,00%	57	0
JOSÉ	50,00%	29,333	0
POPAYÁN	38,50%	33,154	0
CAPITÁN	37,50%	52	0
DON	36,00%	33,56	0
LUGAR	35,70%	20,286	0,009
JUAN	27,30%	54,364	0
NOMBRE	25,00%	32	0
APUNTES	25,00%	23	0,003
SEÑOR	23,50%	103,53	0
PASTO	23,30%	24,4	0,002
PRIMERO	18,20%	29	0
SOCIEDAD	18,20%	50	0
PUEDEN	16,70%	26,667	0,001
GUERRA	15,80%	39,526	0
DOCTOR	15,40%	21,031	0,007
SANTANDER	14,30%	36,286	0
ESPAÑOL	10,00%	33,2	0
CONJETURA	10,00%	26,8	0,001
HUMANA	8,30%	33,333	0
PUEBLOS	8,30%	28	0
HISTORIA	7,70%	20	0,01
LES	5,90%	29,529	0
HIJOS	5,90%	39	0
ESTUDIOS	5,40%	30,676	0
VERDAD	5,00%	24	0,002
OBRA	5,00%	41,6	0
SAÑUDO	4,80%	29,231	0
EXPIACIÓN	3,70%	30,185	0
BOLÍVAR	1,80%	23,421	0,003

Especificidades Exclusivas.

Las unidades léxicas "exclusivas", aquí son las presentes solamente dentro del subconjunto considerado y "no" en otros.

En este caso, se contempla el repetido cruce entre dos filas (formas seleccionadas por mayor frecuencia en el corpus y artículo) y dos columnas (presencias y ausencias de cada palabra); de modo que el test tenga solamente un grado de libertad y un umbral (12.5%).

Así puede observarse de la Tabla 25, las formas exclusivas definidas por el número de ocurrencias de cada una en cada discurso, no incluidas en otro y que conforman la "especificidad exclusiva" del discurso (comúnmente considerando frecuencias > 1 en cada uno).

Tabla 25 Unidades léxicas exclusivas por frecuencia en artículo

	1.	2.	3.	4.	5.	6.	7.	8.
	DCr_ SEO_ IP1	DCr_ SEO_ AC	DCr_ SEO_ IP2	DCc_ JRS_ PG	DCc_ JRS_ PLF1	DCc_ JRS_ PLF2	DCc_ JRS_ PLF3	DCc_ JRS_ PLF4
<i>SOLIDARIDAD</i>			17					
<i>RESPONSABILIDAD</i>			12					
<i>ESPAÑOLES</i>		10						
<i>INDIVIDUO</i>			10					

El proceso de forma similar se puede realizar por pseudónimo, carácter y autor. A continuación se presenta el análisis por autor.

Vocabulario característico por autor

En la tabla 26 se presenta el análisis de vocabulario por autor organizado por frecuencia de palabras (se han incluido las 100 primeras formas sin lematizar, lo mismo que las que implican modales lógicos), presentado el número en cada columna por conteo de palabras que aparecen en el artículo:

Tabla 26 Tabla cruzada tabula presenta conteo de palabras por autor según el orden de frecuencia de aparición.

	SERGIO ELIAS ORTIZ	JOSÉ RAFAEL SAÑUDO	Chi2	P (2- colas)	#¡VALOR!
SAÑUDO	32	72	2,01	0,366	1
SEÑOR		68	40,8	0	1
DOCTOR	31	34	2,881	0,237	1
BOLÍVAR	30	27	5,568	0,062	1
HISTORIA	26	26	3,467	0,177	1
ERA	12	32	1,964	0,375	1
PUES	3	40	17,093	0	1
OBRA	3	37	15,36	0	1
DICE	11	27	1,186	0,553	1
ESTUDIOS	18	19	1,962	0,375	1
CIUDAD		32	19,2	0	1
DECIR	6	25	4,355	0,113	1
PASTO		30	18	0	1
EXPIACIÓN	16	11	5,454	0,065	1
PÁGINA		26	15,6	0	1
TAN	7	18	0,963	0,618	1
CABILDO		24	14,4	0,001	1
NOMBRE	3	21	6,4	0,041	1
MODO	5	18	2,438	0,296	1
JUAN		22	13,2	0,001	1
SOCIEDAD	13	9	4,376	0,112	1
COLOMBIA	6	15	0,714	0,7	1
HECHO	6	15	0,714	0,7	1
NARIÑO	1	20	9,603	0,008	1
HOMBRES	15	5	12	0,002	1
ORTIZ		20	12	0,002	1
PADRE	2	18	6,453	0,04	1
SIENDO	3	17	4,32	0,115	1
GOBERNADOR		19	11,4	0,003	1
GUERRA	13	6	7,751	0,021	1
HIZO	4	15	2,193	0,334	1
MORAL	14	5	10,614	0,005	1
DEBE	4	14	1,793	0,408	1
AÑOS	3	14	2,859	0,239	1
CIERTO	3	14	2,859	0,239	1
HIJOS	11	6	5,369	0,068	1
HOMBRE	10	7	3,298	0,192	1
SOLIDARIDAD	17		28,333	0	1
APUNTES		16	9,6	0,008	1
CAPITÁN		16	9,6	0,008	1
MUCHOS	3	13	2,4	0,301	1
ALGUNOS	2	13	3,738	0,154	1
ERRORES	3	12	1,96	0,375	1
FRANCISCO		15	9	0,011	1

HABER	2	13	3,738	0,154	1
PALABRAS	3	12	1,96	0,375	1
TENÍA	2	13	3,738	0,154	1
DERECHO	11	3	10,076	0,006	1
ET	10	4	6,876	0,032	1
HACER	3	11	1,543	0,462	1
HUMANIDAD	12	2	13,886	0,001	1
LUGAR	3	11	1,543	0,462	1
MUERTE	10	4	6,876	0,032	1
SANTANDER		14	8,4	0,015	1
SIGUIENTE	1	13	5,505	0,064	1
SIQUIERA	3	11	1,543	0,462	1
DIJO	2	11	2,713	0,258	1
DIO		13	7,8	0,02	1
INDIO	2	11	2,713	0,258	1
NÚMERO		13	7,8	0,02	1
PECADOS	13		21,667	0	1
POPAYÁN		13	7,8	0,02	1
SAN	2	11	2,713	0,258	1
TODA	9	4	5,585	0,061	1
ESCRITO	1	11	4,356	0,113	1
HECHOS	3	9	0,8	0,67	1
HUBO	1	11	4,356	0,113	1
HUMANA	9	3	7,2	0,027	1
INDEPENDENCIA	6	6	0,8	0,67	1
JOSÉ		12	7,2	0,027	1
LIBERTADOR	11	1	15,022	0,001	1
PALABRA	3	9	0,8	0,67	1
PESOS		12	7,2	0,027	1
PUEBLOS	10	2	10,756	0,005	1
PUEDEN	10	2	10,756	0,005	1
RESPONSABILIDAD	12		20	0	1
SOLDADOS	1	11	4,356	0,113	1
VARIOS		12	7,2	0,027	1
ALGUNO	1	10	3,788	0,15	1
BOLETÍN		11	6,6	0,037	1
DEPARTAMENTO		11	6,6	0,037	1
DICHO	1	10	3,788	0,15	1
ERROR	2	9	1,752	0,417	1
JEFES	10	1	13,388	0,001	1
PARECE	7	4	3,206	0,201	1
PRIMERO	1	10	3,788	0,15	1
QUITO		11	6,6	0,037	1
REDENCIÓN	7	4	3,206	0,201	1
SEÑORES	1	10	3,788	0,15	1
SISTEMA	10	1	13,388	0,001	1
ALGUNA	1	9	3,227	0,199	1
AÑO		10	6	0,05	1

CONJETURA		10	6	0,05	1
DA	1	9	3,227	0,199	1
ESPAÑOL	7	3	4,507	0,105	1
ESPAÑOLES	10		16,667	0	1
GOBIERNO	8	2	7,707	0,021	1
HE	1	9	3,227	0,199	1
INDIVIDUOS	10		16,667	0	1
JULIO	2	8	1,307	0,52	1
LETRAS	1	9	3,227	0,199	1
MANUEL	1	9	3,227	0,199	1
NATURAL	6	4	2,16	0,34	1
OBISPO		10	6	0,05	1
PRIMER	2	8	1,307	0,52	1
PUENTE		10	6	0,05	1
RAZÓN	2	8	1,307	0,52	1
TODAS	6	4	2,16	0,34	1
TRATADO		10	6	0,05	1
TUVO	2	8	1,307	0,52	1
NI	16	25	0,041	0,98	0
SINO	16	24	0,107	0,948	0
LES	13	21	0,008	0,996	0
VIDA	14	20	0,196	0,907	0
DON	9	16	0,024	0,988	0
DIOS	9	12	0,257	0,879	0
VERDAD	8	12	0,053	0,974	0
ENTONCES	7	9	0,267	0,875	0
MISMA	5	9	0,019	0,991	0
PATRIA	5	9	0,019	0,991	0
TIENEN	5	9	0,019	0,991	0
CADA	4	9	0,251	0,882	0
COSA	6	7	0,415	0,812	0
HISTORIADOR	5	8	0,005	0,997	0
PRINCIPIO	6	7	0,415	0,812	0
PUEBLO	5	7	0,089	0,957	0
SEA	4	8	0,089	0,957	0
FIN	5	6	0,297	0,862	0
JAMÁS	3	8	0,491	0,782	0
MADRE	3	8	0,491	0,782	0
PÚBLICA	4	7	0,006	0,997	0
ESPAÑA	3	7	0,24	0,887	0
OTRO	5	5	0,667	0,717	0
SABIO	4	6	0,027	0,987	0
TANTO	5	5	0,667	0,717	0

En la tabla 26 se incluyó una columna al final, el valor de uno significa que el cálculo de chi cuadrado es superior a p, en caso contrario el valor es 0. Se identificaron las características “relevantes” del vocabulario de cada autor con

validación estadística superior al 95% de certeza ($\alpha=0.05$). Como quiera que en el cálculo del chi-cuadrado en el ejemplo obtuvimos un valor en cada fila, que supera al valor para $\alpha=0.005$, podremos concluir que las variables no son independientes, sino que están asociadas en los casos que el valor del chi cuadrado es superior a p (última columna valor 1). Por lo tanto, a la vista de los resultados, rechazamos la hipótesis nula (H_0) y aceptamos la hipótesis alternativa (H_a) como probablemente cierta.

H_0 : No hay asociación entre las variables (en el ejemplo, los autores entre sí por su vocabulario son independientes, no están asociados).

Y la hipótesis alternativa:

H_a : Sí hay asociación entre las variables, es decir, los autores por su vocabulario están asociados.

En los casos que en la última columna aparece el valor 0 se aceptaría para ellos la Hipòtesis nula. Esto sería para las palabras NI, SINO, LES, VIDA, DON, DIOS, VERDAD, ENTONCES, MISMA, PATRIA, TIENEN, CADA, COSA, HISTORIADOR, PRINCIPIO, PUEBLO, SEA, FIN, JAMÁS, MADRE, PÚBLICA, ESPAÑA, OTRO, SABIO, TANTO.

Los valores de conteo de palabra ordenados por frecuencia en cada artículo, presentados en la tabla 25 se grafican a continuación (ilustración 16):

Frecuencia total para
AUTOR (Porcentaje total
de palabras)

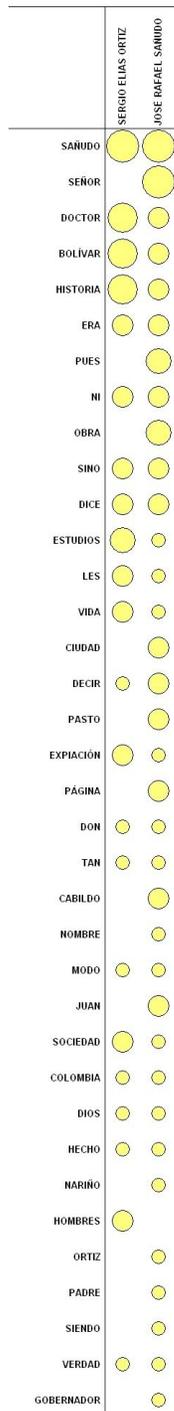


Ilustración 16 Gráfico de burbujas frecuencia de formas por autor

La Clasificación del vocabulario de los discursos por autor permite obtener tipologías de los mismos en las que la progresión del vocabulario se produce en función del motivo y temáticas del artículo.

Palabras típicas y específicas de cada autor.

Con base en la tabla 26 se elaboró la tabla 27 para establecer un listado ordenado con las unidades lexicales (palabras) típicas y exclusivas de cada artículo, en las cuales se cumple la Hipótesis alterna.

Las unidades lexicales "típicas" son definidas por exceso (sobre utilización) según un criterio estadístico que requiere la aplicación del Test de Chi cuadrado. Así puede observarse en la Tabla 4.26, las formas típicas definidas por valores altos del CHI2 y que conforman la "especificidad" del discurso de cada autor (considerando valores de $CHI2 > 20$ en cada uno, se ven las formas más específicas).

Tabla 27 Tabla cruzada presenta unidades lexicales típicas por autor

	SERGIO ELIAS ORTIZ	JOSÉ RAFAEL SAÑUDO	Chi2	P (2-colas)
<i>SEÑOR</i>		68	40,8	0
<i>SOLIDARIDAD</i>	17		28,333	0
<i>PECADOS</i>	13		21,667	0
<i>RESPONSABILIDAD</i>	12		20	0

Se esperaría que las palabras típicas en las réplicas de Sergio Elías Ortiz, estuviesen relacionadas con la impugnación a la concepción de historia de Sañudo (el problema de la expiación) sin embargo el de mayor frecuencia es "solidaridad", como si la inquietud fuese de carácter moral por la existencia o no de solidaridad en los pecados de los líderes y la responsabilidad de los ciudadanos frente a esto; entonces las inquietudes están enmarcadas en las consecuencias de aceptar como ciertos los presupuestos de Sañudo. Por su parte la forma "señor" típica de Sañudo, respondería a la necesidad de plantear el debate en condiciones de urbanidad con un trato formal, puesto que como consecuencia de sus Estudios, la reacción frente a él, ha implicado la declaración como antipatriota y el peligro de ser atacado desconociéndole su derecho a libertad de opinión y conciencia.

Especificidades Exclusivas.

Las unidades léxicas "exclusivas", aquí son las presentes solamente dentro del subconjunto considerado y "no" en otros.

En este caso, se contempla el repetido cruce entre dos filas (formas seleccionadas por mayor frecuencia en el corpus y artículo) y dos columnas (presencias y ausencias de cada palabra).

Así puede observarse de la Tabla 28, las formas exclusivas definidas por el número de ocurrencias de cada una en cada discurso, no incluidas en otro y que conforman la “especificidad exclusiva” del discurso (comúnmente considerando frecuencias > 1 en cada uno).

Tabla 28 Unidades léxicas exclusivas de José Rafael Sañudo ordenadas por frecuencia

	JOSÉ RAFAEL SAÑUDO	Chi2	P (2- colas)
<i>SEÑOR</i>	68	40,8	0
<i>CIUDAD</i>	32	19,2	0
<i>PASTO</i>	30	18	0
<i>PÁGINA</i>	26	15,6	0
<i>CABILDO</i>	24	14,4	0,001
<i>JUAN</i>	22	13,2	0,001
<i>ORTIZ</i>	20	12	0,002
<i>GOBERNADOR</i>	19	11,4	0,003
<i>APUNTES</i>	16	9,6	0,008
<i>CAPITÁN</i>	16	9,6	0,008
<i>FRANCISCO</i>	15	9	0,011
<i>SANTANDER</i>	14	8,4	0,015
<i>DIO</i>	13	7,8	0,02
<i>NÚMERO</i>	13	7,8	0,02
<i>POPAYÁN</i>	13	7,8	0,02
<i>JOSÉ</i>	12	7,2	0,027
<i>PESOS</i>	12	7,2	0,027
<i>VARIOS</i>	12	7,2	0,027
<i>BOLETÍN</i>	11	6,6	0,037
<i>DEPARTAMENTO</i>	11	6,6	0,037
<i>QUITO</i>	11	6,6	0,037
<i>AÑO</i>	10	6	0,05
<i>CONJETURA</i>	10	6	0,05
<i>OBISPO</i>	10	6	0,05
<i>PUENTE</i>	10	6	0,05
<i>TRATADO</i>	10	6	0,05

Tabla 29 Unidades léxicas exclusivas de Sergio Elías Ortiz ordenadas por frecuencia

	SERGIO ELIAS ORTIZ	Chi2	P (2- colas)
<i>SOLIDARIDAD</i>	17	28,333	0
<i>PECADOS</i>	13	21,667	0
<i>RESPONSABILIDAD</i>	12	20	0
<i>ESPAÑOLES</i>	10	16,667	0
<i>INDIVIDUOS</i>	10	16,667	0

Asociación de palabras claves del corpus.

Mediante el procedimiento de asociación de lemas, elegida una palabra clave determinada se identifican los lemas asociados a la misma, con la presencia de las ocurrencias y co- ocurrencias, de la manera siguiente:

1. Se realiza la selección mediante el índice de asociación “coeficiente de Jaccard (tipo co-ocurrencia de palabra primer orden)”.
2. Se compone un listado de los lemas asociados a esa palabra clave, con los índices.

Ilustración 17. Lista de las asociaciones del lema “BOLIVAR”.

CASO #		PALABRAS CLAVE		FILE	CARACTER	AUTOR	SEUDONIMO
1	El más sabio padece errores y el más justo practica	Bolívar	, porque las pasiones ofuscan el entendimiento y	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	En la misma introducción, digna por su vuelo de una	Bolívar	, que sería una revelación. Dios, en su designio	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Y decimos que el doctor Sañudo ha muerto de	Bolívar	; y porque el insultador del Padre de la Patria, debe la	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Después de intentar rápida e inútilmente la refutación	Bolívar	, funda la historia en una base de rara explicación, en	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Y donde se nota la parcialidad del historiador	Bolívar	, es en sostener su curioso principio de explicación,	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Y donde se nota la parcialidad del historiador	Bolívar	—debería explicar. Esto, que no afirmamos nosotros, es	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Métodos un instante en el expiatorio sistema del	Bolívar	un formidable dato histórico de la ley de la explicación?	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	Y sobre la base de que	Bolívar	fue un hombre sujeto como todos a las pasiones, sin	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
1	De suerte que no es	Bolívar	la causa de las penas, de los sacrificios, ni de los	1. DCr_SEO_IP1	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
2	Que la España sepa una vez por todas, que los	Bolívar	el héroe más grande de la América, la Historia no lo	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que la España sepa una vez por todas, que los	Bolívar	su más noble contendor—que no fue el virrey	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que la España sepa una vez por todas, que los	Bolívar	que el inca. Que en Bolívar brilló la justicia el mismo	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que la España sepa una vez por todas, que los	Bolívar	brilló la justicia el mismo Unamuno, español, lo va a	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que la España sepa una vez por todas, que los	Bolívar	. Fusiló a propios y a extraños pero jamás con	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que la España sepa una vez por todas, que los	Bolívar	fusiló inmediatamente a ochocientos prisioneros.	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	, tenían pleno derecho al Gobierno de sus pueblos; y	Bolívar	lo hace aprehender, juzgar y fusilar. Lo propio hizo	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que	Bolívar	fue cruel: lo contrario dice Blanco Fombona de la	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que Bolívar fue cruel: lo contrario dice Blanco	Bolívar	y el éxito lo hacía magnánimo.” Que el Libertador no	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que Bolívar fue cruel: lo contrario dice Blanco	Bolívar	los sorprendía concentrando una fuerza superior	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Que Bolívar fue cruel: lo contrario dice Blanco	Bolívar	fue el autor de la guerra a muerte? Allá va la historia	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Seguimos con que	Bolívar	aprisionó a Miranda. La traición fue la estrella de	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Seguimos con que Bolívar aprisionó a Miranda. La	Bolívar	, y con el conocimiento que tenía de la traición de	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Seguimos con que Bolívar aprisionó a Miranda. La	Bolívar	fue el mayor revolucionario del país? No es cierto.	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
2	Las guerras consideradas en particular y en el	Bolívar	, no haya brillado la imparcialidad al hacer el recuento	2. DCr_SEO_AC	REPLICA	SERGIO ELIAS ORTIZ	ALEXO CORFU
3	A esta teoría, que ahora podría denominarse del	Bolívar	objeciones infantes que después aplasta con	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
3	Para un espíritu resbaladizo, sin firmeza filosófica ni	Bolívar	, afirmadora en el fondo de las ideas, y más	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
3	Para un espíritu resbaladizo, sin firmeza filosófica ni	Bolívar	que desaparecieran las exageraciones partidistas	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
3	Empeñados los Estudios en lucir su sistema	Bolívar	en todas las desgracias de Colombia,	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
3	Empeñados los Estudios en lucir su sistema	Bolívar	.	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
3	Nosotros—en cambio—vernos en las páginas	Bolívar	, solidario de la obra de libertad que ha purificado y	3. DCr_SEO_IP2	REPLICA	SERGIO ELIAS ORTIZ	UN INDIIO PATRIOTA
4	Dió Régulo una regla para analizar esos Estudios, de	Bolívar	, perjurios, falsificaciones y asesinatos y llamé	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	Ni la verdadera crítica histórica sufre que sólo se	Bolívar	en todos sus crímenes, aún en los más probados,	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	Nosotros, que siempre hemos tenido al señor	Bolívar	; pues decididos Indio.	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	¿Tienes conciencia de que queriendo el señor	Bolívar	, para que se enojen los directores de Renacimiento?	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	¿Tienes conciencia de que queriendo el señor	Bolívar	y no la historia de un pueblo, de fugistar, v. g., a doña	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	Hasta llegó a asegurar que había muerto sin	Bolívar	, se alcan más bien al caballero Nariño, o al	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
4	Lo que mtejamos al señor Sañudo, es su	Bolívar	en Bombona había dado a sus abuelos una paliza de	4. DCc_JRS_PG	CONTRARREPLIC	JOSE RAFAEL SANUDO	PROSPERO GALLO
5	Pero el colmo de la desestima del paisano, el colmo	Bolívar	; y lo que es más extraño siendo parte de la fácil	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	Pero el colmo de la desestima del paisano, el colmo	Bolívar	con ofuscamiento de procesos colombianos, muy	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	Pero el colmo de la desestima del paisano, el colmo	Bolívar	. No es honrosa para un pastuso y sobre todo para	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	Juan Bautista Ortiz del Corral, quien dijo no haberla	Bolívar	volviese enemigo de Santander, los últimos cuatro	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	que con menor fortuna, tales como: Nariño,	Bolívar	debió fusilarle, pues le hacen aparecer como un	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	in par injusticia; y apenas vista crueldad; y por último,	Bolívar	, como no de talla de héroe; y no puedo menos de	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	Oí a Sañudo decir a un respetable y simpático	Bolívar	. Nariño y los demás citados que con el sabio Caldas	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	bles entendieron en ella. Ni siquiera han podido	Bolívar	ejecutado cuando unió a la actual Colombia,	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
5	a separada de ella desde la Colonia y acostumbrada	Bolívar	sostiene, como lo haría Nietzsche al decir que el	5. DCc_JRS_PLF1	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Hay quienes encuentran buena la Obra en sí, pero	Bolívar	tuviera, según él opina, arrojos como Népoléon en el	5. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Hay quienes encuentran buena la Obra en sí, pero	Bolívar	que en carta a Santander del 21 de octubre de 1825	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Pero dejemos a un lado estos recuerdos, digamos	Bolívar	, para poder imitar a Iturbide, que se coronó como	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Tiene Sañudo disculpa en haber pasado en silencio	Bolívar	y su vida íntima, pues el ámbito de unos estudios	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Tiene Sañudo disculpa en haber pasado en silencio	Bolívar	como se propuso, sino también la privada, porque	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Tiene Sañudo disculpa en haber pasado en silencio	Bolívar	, es cierto que poco podía escribir Sañudo, porque un	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Por el consiguiente tengo por poco cuerdo al que	Bolívar	. Estas debieron de ser verbales, hasta hoy la historia	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	Por el consiguiente tengo por poco cuerdo al que	Bolívar	habría castigado semejante crimen cometido casi en	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	10 Ahora los quiterios tratan de alzar una estatua a	Bolívar	. Dios les valga por generosos! La Biblia aconseja,	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO
7	10 Ahora los quiterios tratan de alzar una estatua a	Bolívar	puso como chupa de dómíne a los quiterios, en	7. DCc_JRS_PLF3	CONTRARREPLIC	JOSE RAFAEL SANUDO	PATRICIO LEAL Y FRANCO

3. Se construye la tabla que visualiza el lema elegido y la distribución de las palabras seleccionadas más asociadas al mismo.

Finalmente se describen todos los contextos elementales donde empareja con la palabra elegida. Mediante este procedimiento de asociación de lemas, se considera la proximidad a la palabra clave, de otra asociada escogida y se describen todos los contextos elementales donde empareja con la palabra seleccionada.

Tabla 30 proximidad lemas Bolivar e hiatoria con el resto de formas

OBJETIVO	PALABRAS CLAVE	CO-OCURRE	NO	AUSENTE	Coficiente	FUERZA
BOLÍVAR	VIDA	13	11	17	0,317
BOLÍVAR	NI	13	13	17	0,302
HISTORIA	EXPIACIÓN	12	6	26	0,273
BOLÍVAR	COLOMBIA	10	7	20	0,27
BOLÍVAR	LIBERTADOR	8	2	22	0,25
BOLÍVAR	SAÑUDO	19	47	11	0,247
BOLÍVAR	OBRA	10	11	20	0,244
HISTORIA	SAÑUDO	19	47	19	0,224
HISTORIA	DICE	12	17	26	0,218
HISTORIA	DOCTOR	15	31	23	0,217
HISTORIA	VIDA	11	13	27	0,216
BOLÍVAR	SISTEMA	7	3	23	0,212
BOLÍVAR	HIZO	8	8	22	0,211
BOLÍVAR	AÑOS	8	8	22	0,211
BOLÍVAR	TODA	7	5	23	0,2
BOLÍVAR	DON	7	6	23	0,194
HISTORIA	ERA	11	19	27	0,193
BOLÍVAR	DOCTOR	12	34	18	0,188
BOLÍVAR	ESTUDIOS	9	19	21	0,184
BOLÍVAR	PÚBLICA	6	3	24	0,182
BOLÍVAR	DICE	9	20	21	0,18
BOLÍVAR	DEBE	7	9	23	0,179
BOLÍVAR	HOMBRES	7	10	23	0,175
HISTORIA	SISTEMA	7	3	31	0,171
BOLÍVAR	HISTORIADOR	6	5	24	0,171
BOLÍVAR	DERECHO	6	5	24	0,171
BOLÍVAR	VERDAD	7	11	23	0,171
BOLÍVAR	EXPIACIÓN	7	11	23	0,171
HISTORIA	LIBERTADOR	7	3	31	0,171
BOLÍVAR	PUES	9	24	21	0,167
BOLÍVAR	SINO	9	24	21	0,167
HISTORIA	HUMANIDAD	7	4	31	0,167
BOLÍVAR	HECHOS	6	6	24	0,167
HISTORIA	PUES	10	23	28	0,164
HISTORIA	NI	9	17	29	0,164
HISTORIA	SINO	10	23	28	0,164
BOLÍVAR	TENÍA	6	7	24	0,162
BOLÍVAR	SEÑOR	9	26	21	0,161
HISTORIA	HACER	7	6	31	0,159
BOLÍVAR	TIENEN	6	8	24	0,158
BOLÍVAR	SIENDO	6	8	24	0,158
HISTORIA	TIENEN	7	7	31	0,156
HISTORIA	ALGUNOS	7	7	31	0,156
HISTORIA	MUCHOS	7	7	31	0,156
BOLÍVAR	ERA	8	22	22	0,154
BOLÍVAR	HOMBRE	6	9	24	0,154
BOLÍVAR	NARIÑO	6	9	24	0,154
BOLÍVAR	ESPAÑA	5	3	25	0,152

BOLÍVAR	GUERRA	5	3	25	0,152	*****
BOLÍVAR	SABIO	5	4	25	0,147	*****
HISTORIA	HOMBRES	7	10	31	0,146	*****
BOLÍVAR	TAN	6	11	24	0,146	*****
HISTORIA	TUVO	6	3	32	0,146	*****
HISTORIA	DECIR	8	17	30	0,145	*****
BOLÍVAR	MUERTE	5	5	25	0,143	*****
BOLÍVAR	SANTANDER	5	5	25	0,143	*****
HISTORIA	HISTORIADOR	6	5	32	0,14	*****
BOLÍVAR	ALGUNO	5	6	25	0,139	*****
BOLÍVAR	PATRIA	5	6	25	0,139	*****
HISTORIA	DIJO	6	6	32	0,136	*****
HISTORIA	OBRA	7	14	31	0,135	*****
HISTORIA	PASTO	7	15	31	0,132	*****
BOLÍVAR	HABER	5	8	25	0,132	*****
BOLÍVAR	HACER	5	8	25	0,132	*****
HISTORIA	HOMBRE	6	9	32	0,128	*****
BOLÍVAR	MUCHOS	5	9	25	0,128	*****
HISTORIA	NARIÑO	6	9	32	0,128	*****
BOLÍVAR	ALGUNOS	5	9	25	0,128	*****
HISTORIA	DEBE	6	10	32	0,125	*****
HISTORIA	LES	6	10	32	0,125	*****
HISTORIA	SEÑOR	8	27	30	0,123	*****
HISTORIA	TAN	6	11	32	0,122	*****
BOLÍVAR	DECIR	6	19	24	0,122	*****
HISTORIA	ESPAÑA	5	3	33	0,122	*****
HISTORIA	ESTUDIOS	7	21	31	0,119	*****
BOLÍVAR	CIERTO	5	12	25	0,119	*****
HISTORIA	TODAS	5	4	33	0,119	*****
HISTORIA	CIUDAD	6	13	32	0,118	*****
BOLÍVAR	PRIMERO	4	4	26	0,118	*****
HISTORIA	DIOS	6	13	32	0,118	*****
HISTORIA	MODO	6	14	32	0,115	*****
BOLÍVAR	DIOS	5	14	25	0,114	*****
HISTORIA	SIQUIERA	5	6	33	0,114	*****
BOLÍVAR	TODAS	4	5	26	0,114	*****
BOLÍVAR	PRINCIPIO	4	5	26	0,114	*****
BOLÍVAR	TUVO	4	5	26	0,114	*****
HISTORIA	TODA	5	7	33	0,111	*****
BOLÍVAR	INDEPENDENCIA	4	6	26	0,111	*****
HISTORIA	HECHOS	5	7	33	0,111	*****
BOLÍVAR	HUBO	4	6	26	0,111	*****
BOLÍVAR	TANTO	4	6	26	0,111	*****
HISTORIA	PALABRAS	5	8	33	0,109	*****
HISTORIA	DON	5	8	33	0,109	*****
HISTORIA	TENÍA	5	8	33	0,109	*****
BOLÍVAR	SIQUIERA	4	7	26	0,108	*****
HISTORIA	SIENDO	5	9	33	0,106	*****
BOLÍVAR	DIJO	4	8	26	0,105	*****
HISTORIA	REDENCIÓN	4	1	34	0,103	*****
HISTORIA	AÑOS	5	11	33	0,102	*****

HISTORIA	VERDAD	5	13	33	0,098	*****
BOLÍVAR	LES	4	12	26	0,095	*****
HISTORIA	SOLDADOS	4	4	34	0,095	*****
BOLÍVAR	ESPAÑOL	3	2	27	0,094	*****
BOLÍVAR	HECHO	4	13	26	0,093	*****
HISTORIA	SABIO	4	5	34	0,093	*****
HISTORIA	INDEPENDENCIA	4	6	34	0,091	*****
HISTORIA	PUEDEN	4	6	34	0,091	*****
HISTORIA	HUMANA	4	6	34	0,091	*****
HISTORIA	PUEBLO	4	6	34	0,091	*****
HISTORIA	FIN	4	6	34	0,091	*****
HISTORIA	SANTANDER	4	6	34	0,091	*****
HISTORIA	JAMÁS	4	6	34	0,091	*****
HISTORIA	DICHO	4	7	34	0,089	****
HISTORIA	CABILDO	4	7	34	0,089	****
HISTORIA	DIO	4	7	34	0,089	****
HISTORIA	ESCRITO	4	7	34	0,089	****
BOLÍVAR	MANUEL	3	4	27	0,088	****
BOLÍVAR	MODO	4	16	26	0,087	****
HISTORIA	ERRORES	4	8	34	0,087	****
HISTORIA	APUNTES	4	9	34	0,085	****
HISTORIA	HABER	4	9	34	0,085	****
BOLÍVAR	MADRE	3	6	27	0,083	****
BOLÍVAR	VARIOS	3	6	27	0,083	****
HISTORIA	SOCIEDAD	4	10	34	0,083	****
BOLÍVAR	LETRAS	3	6	27	0,083	****
HISTORIA	ENTONCES	4	10	34	0,083	****
BOLÍVAR	JULIO	3	6	27	0,083	****
BOLÍVAR	JAMÁS	3	7	27	0,081	****
BOLÍVAR	PUEBLO	3	7	27	0,081	****
BOLÍVAR	NATURAL	3	7	27	0,081	****
BOLÍVAR	PARECE	3	7	27	0,081	****
BOLÍVAR	INDIO	3	7	27	0,081	****
HISTORIA	HIZO	4	12	34	0,08	****
BOLÍVAR	DICHO	3	8	27	0,079	****
BOLÍVAR	SAN	3	8	27	0,079	****
BOLÍVAR	ESCRITO	3	8	27	0,079	****
HISTORIA	CIERTO	4	13	34	0,078	****
HISTORIA	NOMBRE	4	13	34	0,078	****
BOLÍVAR	SIGUIENTE	3	9	27	0,077	****
BOLÍVAR	ERRORES	3	9	27	0,077	****
HISTORIA	ET	3	1	35	0,077	****
BOLÍVAR	JUAN	3	10	27	0,075	****
BOLÍVAR	ENTONCES	3	11	27	0,073	****
BOLÍVAR	PADRE	3	11	27	0,073	****
HISTORIA	GOBIERNO	3	4	35	0,071	****
BOLÍVAR	MORAL	3	12	27	0,071	****
HISTORIA	GUERRA	3	5	35	0,07	***
HISTORIA	HE	3	5	35	0,07	***
HISTORIA	CONJETURA	3	5	35	0,07	***
HISTORIA	PRINCIPIO	3	6	35	0,068	***

HISTORIA	PUEBLOS	3	6	35	0,068	***
HISTORIA	PÚBLICA	3	6	35	0,068	***
BOLÍVAR	NOMBRE	3	14	27	0,068	***
HISTORIA	HIJOS	3	7	35	0,067	***
HISTORIA	OTRO	3	7	35	0,067	***
HISTORIA	CAPITÁN	3	7	35	0,067	***
BOLÍVAR	CIUDAD	3	16	27	0,065	***
HISTORIA	ORTIZ	3	8	35	0,065	***
HISTORIA	PATRIA	3	8	35	0,065	***
HISTORIA	DERECHO	3	8	35	0,065	***
HISTORIA	MISMA	3	9	35	0,064	***
HISTORIA	SIGUIENTE	3	9	35	0,064	***
BOLÍVAR	ESPAÑOLES	2	2	28	0,063	***
BOLÍVAR	ET	2	2	28	0,063	***
BOLÍVAR	PASTO	3	19	27	0,061	***
HISTORIA	PÁGINA	3	11	35	0,061	***
BOLÍVAR	JEFES	2	4	28	0,059	***
HISTORIA	COLOMBIA	3	14	35	0,058	***
HISTORIA	HECHO	3	14	35	0,058	***
BOLÍVAR	PECADOS	2	5	28	0,057	***
BOLÍVAR	JOSÉ	2	5	28	0,057	***
BOLÍVAR	GOBIERNO	2	5	28	0,057	***
BOLÍVAR	HE	2	6	28	0,056	***
BOLÍVAR	AÑO	2	6	28	0,056	***
BOLÍVAR	ERROR	2	6	28	0,056	***
BOLÍVAR	ALGUNA	2	6	28	0,056	***
BOLÍVAR	PUEBLOS	2	7	28	0,054	***
BOLÍVAR	SEÑORES	2	7	28	0,054	***
BOLÍVAR	SEA	2	7	28	0,054	***
BOLÍVAR	HIJOS	2	8	28	0,053	***
BOLÍVAR	OTRO	2	8	28	0,053	***
BOLÍVAR	PUEDEN	2	8	28	0,053	***
BOLÍVAR	DIO	2	9	28	0,051	***
BOLÍVAR	NÚMERO	2	9	28	0,051	***
BOLÍVAR	ORTIZ	2	9	28	0,051	***
HISTORIA	ESPAÑOL	2	3	36	0,049	**
HISTORIA	PESOS	2	3	36	0,049	**
HISTORIA	DEPARTAMENTO	2	3	36	0,049	**
BOLÍVAR	SOCIEDAD	2	12	28	0,048	**
BOLÍVAR	PÁGINA	2	12	28	0,048	**
HISTORIA	JOSÉ	2	5	36	0,047	**
HISTORIA	MANUEL	2	5	36	0,047	**
HISTORIA	PECADOS	2	5	36	0,047	**
HISTORIA	ALGUNA	2	6	36	0,045	**
HISTORIA	PRIMER	2	6	36	0,045	**
HISTORIA	PRIMERO	2	6	36	0,045	**
HISTORIA	AÑO	2	6	36	0,045	**
HISTORIA	LUGAR	2	7	36	0,044	**
HISTORIA	CADA	2	7	36	0,044	**
HISTORIA	SEÑORES	2	7	36	0,044	**
HISTORIA	JULIO	2	7	36	0,044	**

HISTORIA	POPAYÁN	2	7	36	0,044	••
HISTORIA	MADRE	2	7	36	0,044	••
HISTORIA	RAZÓN	2	8	36	0,043	••
HISTORIA	INDIO	2	8	36	0,043	••
HISTORIA	ALGUNO	2	9	36	0,043	••
HISTORIA	PARECE	2	8	36	0,043	••
HISTORIA	NÚMERO	2	9	36	0,043	••
HISTORIA	MUERTE	2	8	36	0,043	••
HISTORIA	JUAN	2	11	36	0,041	••
HISTORIA	MORAL	2	13	36	0,039	••
BOLÍVAR	PUENTE	1	2	29	0,031	••
BOLÍVAR	TRATADO	1	4	29	0,029	•
BOLÍVAR	PALABRA	1	6	29	0,028	•
BOLÍVAR	SOLDADOS	1	7	29	0,027	•
BOLÍVAR	RESPONSABILIDAD	1	7	29	0,027	•
BOLÍVAR	PRIMER	1	7	29	0,027	•
BOLÍVAR	FIN	1	9	29	0,026	•
BOLÍVAR	FRANCISCO	1	9	29	0,026	•
BOLÍVAR	LUGAR	1	8	29	0,026	•
BOLÍVAR	HUMANA	1	9	29	0,026	•
BOLÍVAR	CADA	1	8	29	0,026	•
BOLÍVAR	GOBERNADOR	1	8	29	0,026	•
BOLÍVAR	RAZÓN	1	9	29	0,026	•
BOLÍVAR	COSA	1	10	29	0,025	•
HISTORIA	ESPAÑOLES	1	3	37	0,024	•
BOLÍVAR	MISMA	1	11	29	0,024	•
HISTORIA	INDIVIDUOS	1	4	37	0,024	•
BOLÍVAR	APUNTES	1	12	29	0,024	•
BOLÍVAR	PALABRAS	1	12	29	0,024	•
HISTORIA	PALABRA	1	6	37	0,023	•
HISTORIA	OBISPO	1	6	37	0,023	•
HISTORIA	BOLETÍN	1	8	37	0,022	•
HISTORIA	QUITO	1	7	37	0,022	•
HISTORIA	SOLIDARIDAD	1	8	37	0,022	•
HISTORIA	GOBERNADOR	1	8	37	0,022	•
HISTORIA	DA	1	8	37	0,022	•
HISTORIA	SEA	1	8	37	0,022	•
HISTORIA	VARIOS	1	8	37	0,022	•
HISTORIA	HUBO	1	9	37	0,021	•
HISTORIA	COSA	1	10	37	0,021	•
HISTORIA	SAN	1	10	37	0,021	•
HISTORIA	NATURAL	1	9	37	0,021	•
HISTORIA	TANTO	1	9	37	0,021	•

En la ilustración 18, se han seleccionado los lemas BOLIVAR E HISTORIA”,

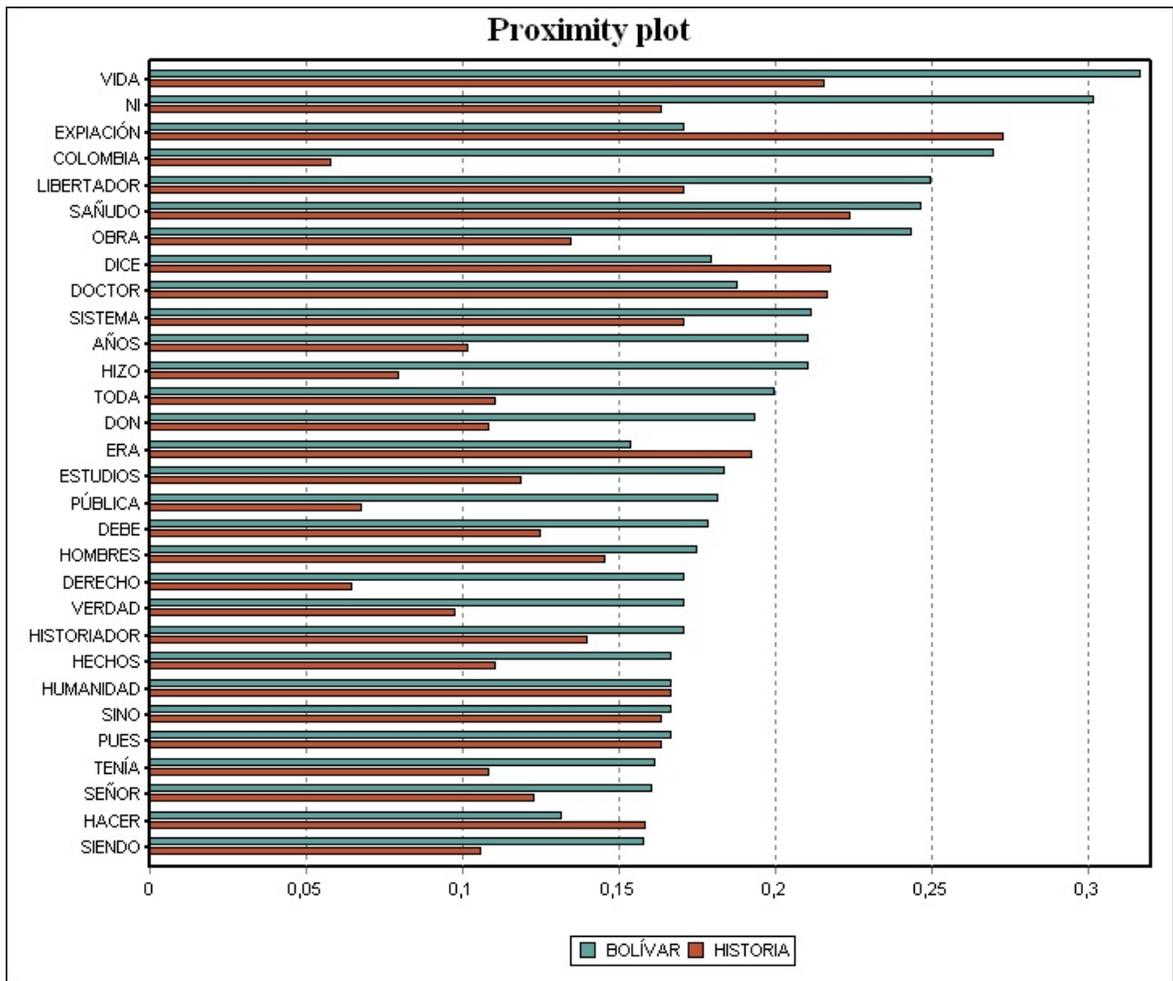


Ilustración 18. Palabras asociadas a la forma “BOLIVAR” e “HISTORIA”

En este proceso se puede recoger y representar el conocimiento en el dominio de palabras claves en los discursos y que conforman una idea cognitiva o una opinión característica.

Las asociaciones se distribuyen alrededor del lema elegido a distancias proporcionales a su grado de asociación dado por el coeficiente de Jaccard.

D. Estudio de la tabla léxica agregada

En primer lugar se analiza la tabla léxica de vocabulario * artículos (tabla léxica transpuesta), para agrupar el vocabulario en función de su distribución en los escritos.

Luego se construye y analiza la tabla léxica agregada de vocabulario * grupos de artículos, donde se agrupan los escritos en columnas en función del carácter, seudónimo y autor.

D.1. Creación de la tabla léxica agregada

Análisis de la tabla léxica (vocabulario * artículos)

Se construye y examina la tabla léxica transpuesta (vocabulario * artículos) de la tabla léxica base, para agrupar las formas según su distribución en los discursos.

D.2. Representación gráfica

Procedimiento realizado:

1. Aplicar el análisis factorial [AFCS] a la tabla transpuesta, que se semeja al análisis factorial de la tabla básica, en razón de la simetría de la misma.
2. Efectuar la clasificación sobre los factores del vocabulario, es decir, agrupar formas que se asocian por mayor ocurrencia (frecuencia) en determinados discursos. Sobre la matriz de factores del vocabulario es aplicado el procedimiento de Clasificación Jerárquica Directa, a través del método de Ward

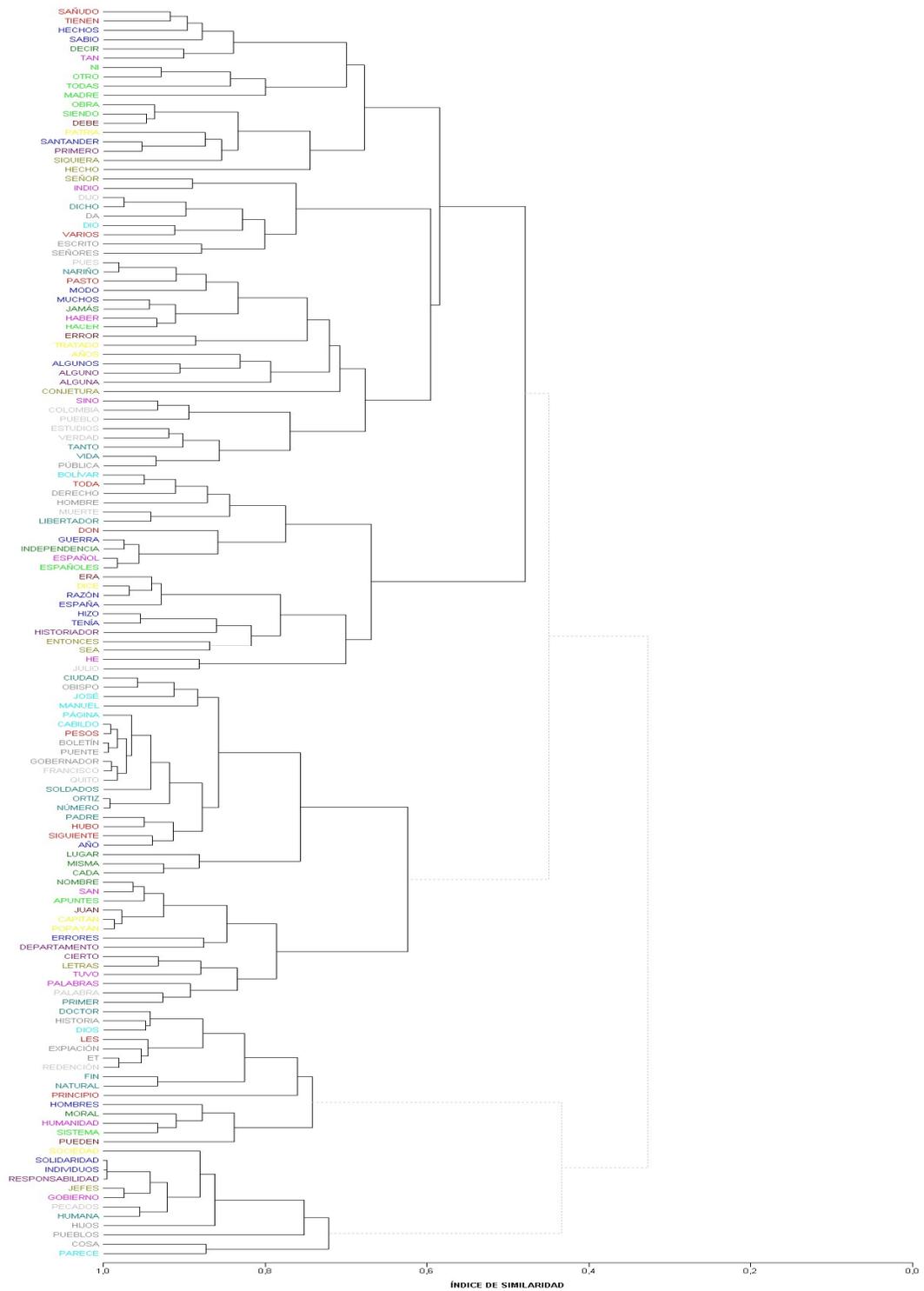


Ilustración 19, visualiza el árbol o Dendograma con las agregaciones efectuadas.

- Cluster 5: apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.
- Cluster 6: Colombia, tratado, departamento, razón, he
- Cluster 7: debe, haber, varios, hijos, padre, error, primer, tuvo, indio.
- Cluster 8: alguno, errores, pues.
- Cluster 9: Bolivar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador.
- Cluster 10: José, tienen, ni, Santander, siquiera, Patria
- Cluster 11: años, cierto, modo, tan, pública
- Cluster 12: algunos, escrito, toda, verdad
- Cluster 13: cosa, tanto, dicho, Dios
- Cluster 14: decir, hecho, otro
- Cluster 15: cada, señores
- Cluster 16: dio, soldados, guerra, pueden
- Cluster 17: España, independencia, Julio, Manuel, español, españoles,
- Cluster 18: muerte, natural
- Cluster 19: fin, letras, palabra, palabras, parece
- Cluster 20: jamás
- Cluster 21: et, les
- Cluster 22: Derecho, gobierno, todas
- Cluster 23: jefes, pecados, pueblos, moral
- Cluster 24: humana
- Cluster 25: responsabilidad, solidaridad, sistema
- Cluster 26: expiación, redención, principio, sabio
- Cluster 27: hombre, hombres, humanidad, individuos, sociedad

Los Cluster 3 y 9 son los de mayor vocabulario, de palabras comunes a la mayoría de los artículos: cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente, Bolivar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador.

El Cluster 23 de menor vocabulario, posee formas más especializadas y únicas para ciertos grupos de discursos, con alto contenido de reproche al proceso de independencia: jefes, pecados, pueblos, moral.

El Cluster 25, contiene vocabulario típico, como se señaló más arriba por frecuencia, de la réplica de Sergio Elías Ortiz: responsabilidad, solidaridad, sistema.

En cambio en el Cluster 26 agrupa un vocabulario especial que formaría parte destacada en la controversia, pero por la frecuencia de términos y la coocurrencia de formas, tiene impacto local: expiación, redención, principio, sabio. Se esperaría que el centro en la réplica la ocupase el cluster 26. En el gráfico factorial

(ilustración 22) aparece el término “entonces” (perteneciente al cluster 5) en el centro, lo cual da indicios de la preocupación de los artículos por mantenerse en una lógica que establece consecuentes. El cluster 5 contiene los términos: apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.

4. Se representan en el plano factorial la distribución de los Cluster.

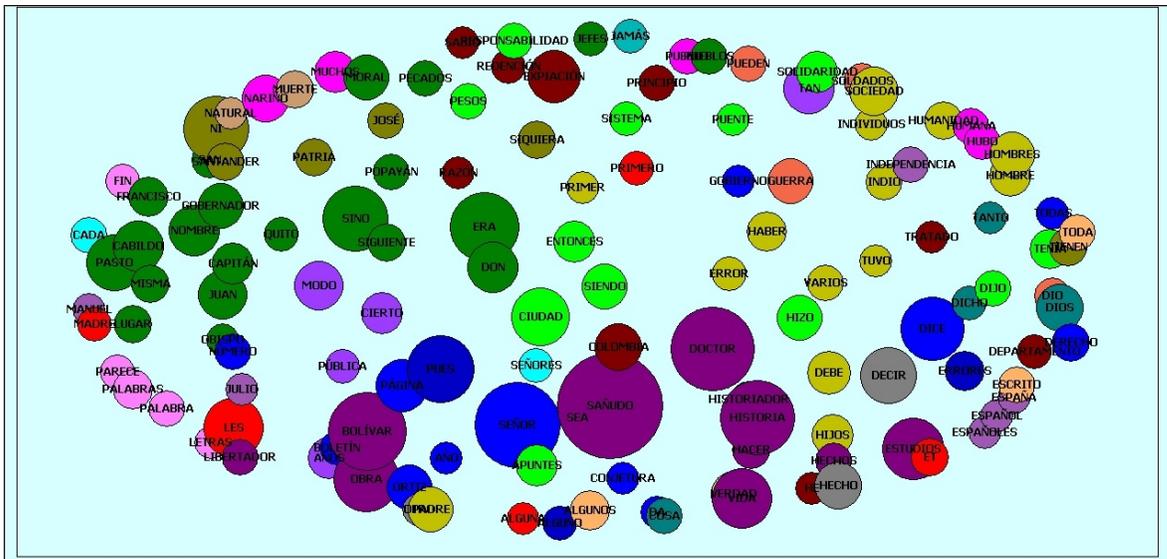


Ilustración 22 Vocabulario por cluster

La ilustración 22, representa, en el plano del vocabulario, la distribución de las diferentes clases, identificándose por colores los distintos grupos de formas localizados en cada grupo.

Hay evidentemente, diferencias de vocabulario en los grupos de discursos clasificados y esto se observa en la representación, identificándose en él los distintos Cluster. La panorámica permite observar globalmente estas diferencias y sus contribuciones están dadas por el tamaño del punto representado, es decir, los grupos que más contribuyen a la riqueza del vocabulario utilizado, son los representados por los puntos de mayor tamaño (el algoritmo utilizado proyecta todos los grupos calculando el tamaño del punto representativo a la contribución de la inercia en el plano principal).

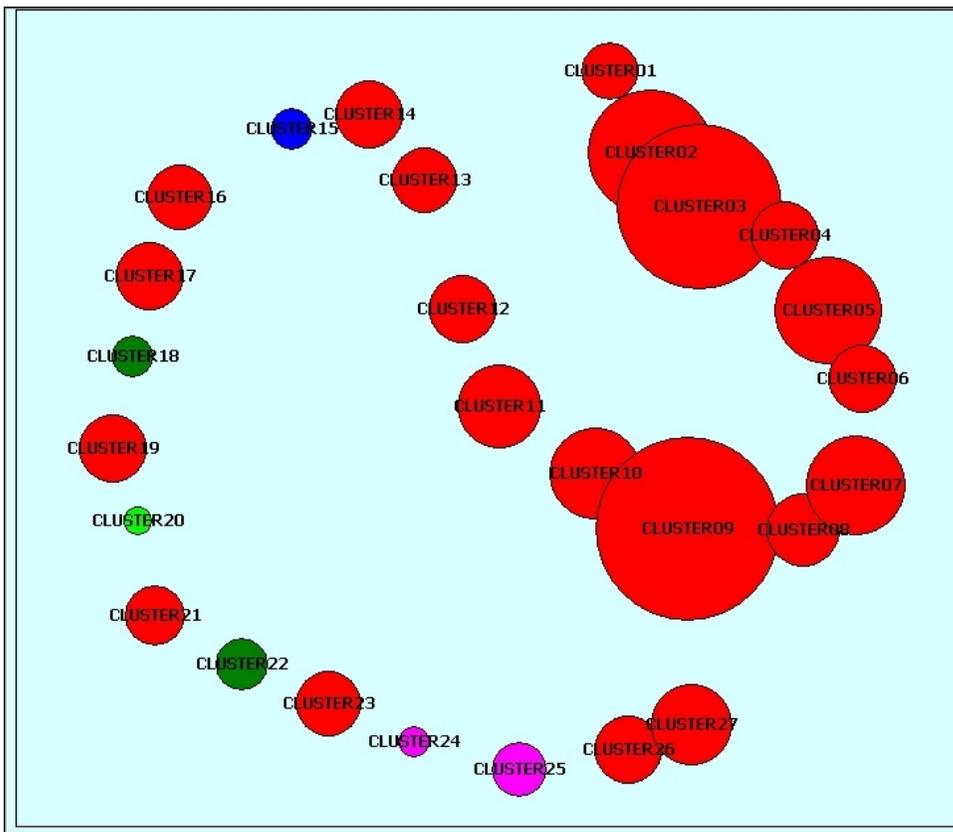


Ilustración 23 Presenta la distribución de los cluster de palabras en el plano

En la Tabla 31, se identifica la frecuencia de las formas de cada clase, presentes en la totalidad de los artículos.

Tabla 31. Frecuencia de palabras agrupadas en clusters presentes en los artículos.

	FRECUENCIA	% MOSTRADO	% PROCESADO	% TOTAL	NO. CASOS	% CASOS	TF • IDF ⁶¹
CLUSTER09	440	17,90%	4,20%	2,00%	8	100,00%	0
CLUSTER03	348	14,20%	3,30%	1,60%	8	100,00%	0
CLUSTER02	206	8,40%	2,00%	0,90%	8	100,00%	0
CLUSTER05	153	6,20%	1,50%	0,70%	8	100,00%	0
CLUSTER07	126	5,10%	1,20%	0,60%	8	100,00%	0
CLUSTER10	109	4,40%	1,00%	0,50%	8	100,00%	0
CLUSTER11	93	3,80%	0,90%	0,40%	8	100,00%	0
CLUSTER27	83	3,40%	0,80%	0,40%	7	87,50%	4,8
CLUSTER08	69	2,80%	0,70%	0,30%	8	100,00%	0
CLUSTER06	62	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER14	62	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER17	62	2,50%	0,60%	0,30%	7	87,50%	3,6
CLUSTER04	61	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER26	61	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER12	60	2,40%	0,60%	0,30%	8	100,00%	0
CLUSTER19	59	2,40%	0,60%	0,30%	8	100,00%	0
CLUSTER16	56	2,30%	0,50%	0,30%	8	100,00%	0
CLUSTER13	55	2,20%	0,50%	0,30%	8	100,00%	0
CLUSTER23	55	2,20%	0,50%	0,30%	6	75,00%	6,9
CLUSTER21	48	2,00%	0,50%	0,20%	6	75,00%	6
CLUSTER01	44	1,80%	0,40%	0,20%	8	100,00%	0
CLUSTER25	40	1,60%	0,40%	0,20%	4	50,00%	12
CLUSTER22	34	1,40%	0,30%	0,20%	7	87,50%	2
CLUSTER15	24	1,00%	0,20%	0,10%	8	100,00%	0
CLUSTER18	24	1,00%	0,20%	0,10%	7	87,50%	1,4
CLUSTER24	12	0,50%	0,10%	0,10%	5	62,50%	2,4
CLUSTER20	11	0,40%	0,10%	0,10%	8	100,00%	0

⁶¹ TF-IDF

Esta medida, propuesta por G. Salton (1989), permite comprobar el peso de un termino (unidad lexical) en un documento (unidad de contexto).

Su fórmula es la siguiente:

$w_{i,j} = tf_{i,j} \times idf_i$ (*Term Frequency x Inverse Document Frequency*)

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

En la que:

$tf_{i,j}$ = número de ocurrencias del termino i en el documento j

idf_i = número de documentos que contienen i

N = número de documentos en una colección (corpus)

El valor $tf_{i,j}$ (Frecuencia del Termino) puede ser normalizada en la manera siguiente:

$$tf_{i,j} = \frac{tf_{i,j}}{\text{Max}(f_{i,j})}$$

en la que $\text{Max}(f_{i,j})$ es la frecuencia máxima de un cualquier termino i en el documento j .

De acuerdo a la tabla 31, los cluster cuyas palabras aparecen con mayor frecuencia en los textos son el 09, 03, 02, al parecer estas agrupaciones serían las que tendrían más relevancia pero el TF IDF nos da otra indicación. Supongamos que en el conjunto de documentos de texto se desea determinar qué documento es más relevante para la consulta de "la historia es redención, no expiación". Una forma sencilla de empezar es mediante la eliminación de documentos que no contengan todas las palabras "la", "historia", "no", "es", "redención" y "expiación", pero esto todavía deja muchos documentos. Para distinguirlos aún más, podríamos contar el número de veces que cada término se produce en cada documento y suma todos juntos; el número de veces que un término aparece en un documento se llama su *frecuencia de término*. Sin embargo, debido a que el término "la" es tan común, esto tiende a incorrectamente hincapié en documentos que utilizan la palabra "la" más frecuentemente, sin dar suficiente peso a los más significativos términos "historia" "expiación" y "redención". También el término "la" no es una buena palabra clave para distinguir términos y documentos pertinentes y no pertinentes. Por el contrario, las palabras "expiación" y "redención" que ocurren rara vez son buena palabras clave para distinguir los documentos pertinentes de los documentos no relevantes. Por lo tanto un factor de *frecuencia inversa del documento* incorporado disminuye el peso de los términos que son muy frecuentes en la colección y aumenta el peso de los términos que ocurren rara vez.

Por ello los cluster más relevantes serían en su orden como se indica en la tabla 32, el cluster 25, le seguiría 23, 21,27,17,24,22,y 18.

Tabla 32 Muestra el orden de relevancia de los términos agrupados en los cluster

	FRECUENCIA	% MOSTRADO	% PROCESADO	% TOTAL	NO. CASOS	% CASOS	TF • IDF
CLUSTER25	40	1,60%	0,40%	0,20%	4	50,00%	12
CLUSTER23	55	2,20%	0,50%	0,30%	6	75,00%	6,9
CLUSTER21	48	2,00%	0,50%	0,20%	6	75,00%	6
CLUSTER27	83	3,40%	0,80%	0,40%	7	87,50%	4,8
CLUSTER17	62	2,50%	0,60%	0,30%	7	87,50%	3,6
CLUSTER24	12	0,50%	0,10%	0,10%	5	62,50%	2,4
CLUSTER22	34	1,40%	0,30%	0,20%	7	87,50%	2
CLUSTER18	24	1,00%	0,20%	0,10%	7	87,50%	1,4
CLUSTER09	440	17,90%	4,20%	2,00%	8	100,00%	0
CLUSTER03	348	14,20%	3,30%	1,60%	8	100,00%	0
CLUSTER02	206	8,40%	2,00%	0,90%	8	100,00%	0
CLUSTER05	153	6,20%	1,50%	0,70%	8	100,00%	0
CLUSTER07	126	5,10%	1,20%	0,60%	8	100,00%	0
CLUSTER10	109	4,40%	1,00%	0,50%	8	100,00%	0
CLUSTER11	93	3,80%	0,90%	0,40%	8	100,00%	0
CLUSTER08	69	2,80%	0,70%	0,30%	8	100,00%	0
CLUSTER06	62	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER14	62	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER04	61	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER26	61	2,50%	0,60%	0,30%	8	100,00%	0
CLUSTER12	60	2,40%	0,60%	0,30%	8	100,00%	0
CLUSTER19	59	2,40%	0,60%	0,30%	8	100,00%	0
CLUSTER16	56	2,30%	0,50%	0,30%	8	100,00%	0
CLUSTER13	55	2,20%	0,50%	0,30%	8	100,00%	0
CLUSTER01	44	1,80%	0,40%	0,20%	8	100,00%	0
CLUSTER15	24	1,00%	0,20%	0,10%	8	100,00%	0
CLUSTER20	11	0,40%	0,10%	0,10%	8	100,00%	0

La tabla 33 muestra la tabla cruzada agrupaciones de palabras por cada uno de los artículos de Sergio Elías Ortiz y José Rafael Sañudo

Tabla 33 Tabla cruzada agrupaciones de palabras y artículos de Sergio Elías Ortiz y José Rafael Sañudo

	1. DCr_ SEO_ IP1	2. DCr_ SEO_ _AC	3. DCr_S EO_IP 2	4. DCc_J RS_P G	5. DCc_J RS_P LF1	6. DCc_J RS_P LF2	7. DCc_J RS_P LF3	8. DCc_J RS_P LF4	Chi2	P (2- colas)
CLUSTER09	74	36	66	56	74	41	68	25	44,909	0
CLUSTER03	8	20	23	18	32	108	26	113	283,034	0
CLUSTER02	3	6	3	44	21	32	19	78	178,466	0
CLUSTER05	1	10	7	14	27	32	17	45	77,745	0
CLUSTER07	3	3	21	25	18	29	16	11	40,73	0
CLUSTER10	9	5	15	7	39	12	7	15	61,202	0
CLUSTER11	6	6	10	13	16	14	17	11	10,484	0,233
CLUSTER27	13	6	41	6	5		7	5	111,795	0
CLUSTER08	3	2	2	8	12	17	14	11	27,348	0,001
CLUSTER06	1	4	4	10	13	10	12	8	16,71	0,033
CLUSTER14	3	2	12	9	13	7	11	5	15,677	0,047
CLUSTER17	2	27		1	10	5	2	15	78,387	0
CLUSTER04	1	3	6	7	9	10	13	12	16,246	0,039
CLUSTER26	23	1	9	17	3	2	5	1	62,148	0
CLUSTER12	5	6	9	7	10	2	14	7	12	0,151
CLUSTER19	4	2	13	9	7	22	1	1	50,153	0
CLUSTER16	4	12	8	5	2	10	3	12	16,286	0,038
CLUSTER13	7	2	12	9	3	3	10	9	14,382	0,072
CLUSTER23	7	4	36		3		4	1	146,745	0
CLUSTER21	23			8	6	3	6	2	65	0
CLUSTER01	2	2	5	4	14	4	4	9	21,091	0,007
CLUSTER25	4	1	34		1				194,8	0
CLUSTER22	2	7	16	1	6		1	1	47,882	0
CLUSTER15	1	1	3	5	2	2	2	8	13,333	0,101
CLUSTER18	8	4	4	1	1	3		3	14,667	0,066
CLUSTER24	1		8	1			1	1	33,333	0
CLUSTER20	1	1	1	1	1	3	2	1	2,818	0,945

Se identificaron las características “relevantes” de cada artículo, con validación estadística superior al 95% de certeza ($\alpha = 0.05$). Como quiera que en el cálculo del chi-cuadrado obtuvimos un valor en cada fila, que supera al valor para $\alpha = 0.005$, podremos concluir que las variables no son independientes, sino que están asociadas teniendo en cuenta que el valor del chi cuadrado es superior a p (un dato más extremo que el que proporciona el test o, equivalentemente, la probabilidad de obtener los datos observados si fuese cierta la hipótesis de independencia). Por lo tanto, a la vista de los resultados, rechazamos la hipótesis nula (H_0) y aceptamos la hipótesis alternativa (H_a) como probablemente cierta.

H_0 : No hay asociación entre las variables (en el ejemplo, los documentos entre sí por su vocabulario agrupado son independientes, no están asociados).

Y la hipótesis alternativa:

H_a: Sí hay asociación entre las variables, es decir, los documentos por su vocabulario agrupado están asociados.

En la ilustración 24, en el plano están identificadas las clases y los documentos asociados a las mismas.

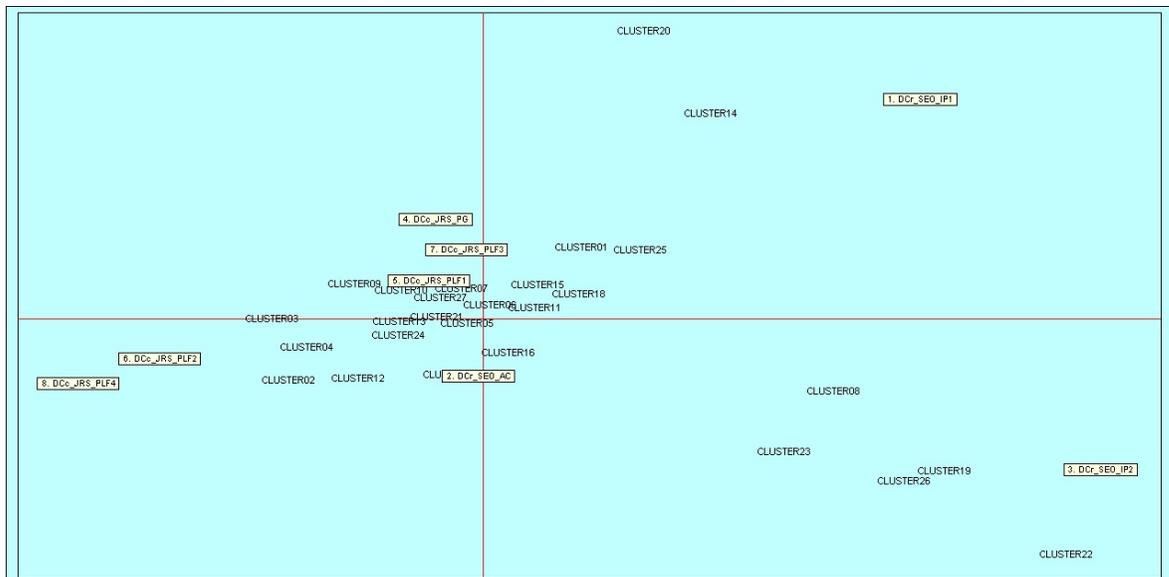


Ilustración 24. Representación del vocabulario agrupado y los artículos de Sergio Elías Ortiz y José Rafael Sañudo.

Para una mejor interpretación es conveniente ayudarse con los procedimientos de especificidades, describiendo y validando estadísticamente los discursos característicos de cada clase.

D.3. Estudio de las especificidades datos complementarios

Descripción y validación estadística de los discursos característicos de cada clase

Se procedió a identificar laa caracterización de los documentos con relación a cada Cluster, descrito en la Tabla 33

Tabla 34 Frecuencia de las agrupaciones de palabras en el primer artículo de Sergio Elías Ortiz firmado con el seudónimo un indio patriota

	1. DCr_SEO_IP1	Chi2	P (2- colas)
CLUSTER09	74	44,909	0
CLUSTER21	23	65	0
CLUSTER26	23	62,148	0
CLUSTER27	13	111,795	0
CLUSTER10	9	61,202	0
CLUSTER03	8	283,034	0
CLUSTER18	8	14,667	0,066
CLUSTER23	7	146,745	0
CLUSTER13	7	14,382	0,072
CLUSTER11	6	10,484	0,233
CLUSTER12	5	12	0,151
CLUSTER25	4	194,8	0
CLUSTER19	4	50,153	0
CLUSTER16	4	16,286	0,038
CLUSTER02	3	178,466	0
CLUSTER07	3	40,73	0
CLUSTER08	3	27,348	0,001
CLUSTER14	3	15,677	0,047
CLUSTER17	2	78,387	0
CLUSTER22	2	47,882	0
CLUSTER01	2	21,091	0,007
CLUSTER05	1	77,745	0
CLUSTER24	1	33,333	0
CLUSTER06	1	16,71	0,033
CLUSTER04	1	16,246	0,039
CLUSTER15	1	13,333	0,101
CLUSTER20	1	2,818	0,945

Tabla 35 Frecuencia de las agrupaciones de palabras en el segundo artículo de Sergio Elías Ortiz firmado con el seudónimo Alexo Corfú

	2. DCr_SEO_AC	Chi2	P (2- colas)
CLUSTER09	36	44,909	0
CLUSTER17	27	78,387	0
CLUSTER03	20	283,034	0
CLUSTER16	12	16,286	0,038
CLUSTER05	10	77,745	0
CLUSTER22	7	47,882	0
CLUSTER27	6	111,795	0
CLUSTER11	6	10,484	0,233
CLUSTER12	6	12	0,151
CLUSTER02	6	178,466	0
CLUSTER10	5	61,202	0
CLUSTER18	4	14,667	0,066
CLUSTER23	4	146,745	0
CLUSTER06	4	16,71	0,033
CLUSTER07	3	40,73	0
CLUSTER04	3	16,246	0,039
CLUSTER13	2	14,382	0,072
CLUSTER19	2	50,153	0
CLUSTER08	2	27,348	0,001
CLUSTER14	2	15,677	0,047
CLUSTER01	2	21,091	0,007
CLUSTER26	1	62,148	0
CLUSTER25	1	194,8	0
CLUSTER15	1	13,333	0,101
CLUSTER20	1	2,818	0,945

Tabla 36 Frecuencia de las agrupaciones de palabras en el tercer artículo de Sergio Elías Ortiz, segundo firmado con el seudónimo un indio patriota

	3. DCr_SEO_IP2	Chi2	P (2- colas)
CLUSTER09	66	44,909	0
CLUSTER27	41	111,795	0
CLUSTER23	36	146,745	0
CLUSTER25	34	194,8	0
CLUSTER03	23	283,034	0
CLUSTER07	21	40,73	0
CLUSTER22	16	47,882	0
CLUSTER10	15	61,202	0
CLUSTER19	13	50,153	0
CLUSTER13	12	14,382	0,072
CLUSTER14	12	15,677	0,047
CLUSTER11	10	10,484	0,233
CLUSTER12	9	12	0,151
CLUSTER26	9	62,148	0
CLUSTER16	8	16,286	0,038
CLUSTER24	8	33,333	0
CLUSTER05	7	77,745	0
CLUSTER04	6	16,246	0,039
CLUSTER01	5	21,091	0,007
CLUSTER18	4	14,667	0,066
CLUSTER06	4	16,71	0,033
CLUSTER02	3	178,466	0
CLUSTER15	3	13,333	0,101
CLUSTER08	2	27,348	0,001
CLUSTER20	1	2,818	0,945

Tabla 37 Frecuencia de las agrupaciones de palabras en la primera contrarréplica de José Rafael Sañudo, firmada con el seudónimo Próspero Gallo

	4. DCc_JRS_PG	Chi2	P (2- colas)
CLUSTER09	56	44,909	0
CLUSTER02	44	178,466	0
CLUSTER07	25	40,73	0
CLUSTER03	18	283,034	0
CLUSTER26	17	62,148	0
CLUSTER05	14	77,745	0
CLUSTER11	13	10,484	0,233
CLUSTER06	10	16,71	0,033
CLUSTER19	9	50,153	0
CLUSTER13	9	14,382	0,072
CLUSTER14	9	15,677	0,047
CLUSTER08	8	27,348	0,001
CLUSTER21	8	65	0
CLUSTER10	7	61,202	0
CLUSTER12	7	12	0,151
CLUSTER04	7	16,246	0,039
CLUSTER27	6	111,795	0
CLUSTER16	5	16,286	0,038
CLUSTER15	5	13,333	0,101
CLUSTER01	4	21,091	0,007
CLUSTER22	1	47,882	0
CLUSTER24	1	33,333	0
CLUSTER18	1	14,667	0,066
CLUSTER20	1	2,818	0,945
CLUSTER17	1	78,387	0

Tabla 38 Frecuencia de las agrupaciones de palabras en la segunda contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco

	5. DCc_JRS_PLF1	Chi2	P (2- colas)
CLUSTER09	74	44,909	0
CLUSTER10	39	61,202	0
CLUSTER03	32	283,034	0
CLUSTER05	27	77,745	0
CLUSTER02	21	178,466	0
CLUSTER07	18	40,73	0
CLUSTER11	16	10,484	0,233
CLUSTER01	14	21,091	0,007
CLUSTER06	13	16,71	0,033
CLUSTER14	13	15,677	0,047
CLUSTER08	12	27,348	0,001
CLUSTER12	10	12	0,151
CLUSTER17	10	78,387	0
CLUSTER04	9	16,246	0,039
CLUSTER19	7	50,153	0
CLUSTER21	6	65	0
CLUSTER22	6	47,882	0
CLUSTER27	5	111,795	0
CLUSTER26	3	62,148	0
CLUSTER13	3	14,382	0,072
CLUSTER23	3	146,745	0
CLUSTER16	2	16,286	0,038
CLUSTER15	2	13,333	0,101
CLUSTER18	1	14,667	0,066
CLUSTER20	1	2,818	0,945
CLUSTER25	1	194,8	0

Tabla 39 Frecuencia de las agrupaciones de palabras en la tercera contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco

	6. DCc_JRS_PLF2	Chi2	P (2- colas)
CLUSTER03	108	283,034	0
CLUSTER09	41	44,909	0
CLUSTER05	32	77,745	0
CLUSTER02	32	178,466	0
CLUSTER07	29	40,73	0
CLUSTER19	22	50,153	0
CLUSTER08	17	27,348	0,001
CLUSTER11	14	10,484	0,233
CLUSTER10	12	61,202	0
CLUSTER06	10	16,71	0,033
CLUSTER04	10	16,246	0,039
CLUSTER16	10	16,286	0,038
CLUSTER14	7	15,677	0,047
CLUSTER17	5	78,387	0
CLUSTER01	4	21,091	0,007
CLUSTER21	3	65	0
CLUSTER13	3	14,382	0,072
CLUSTER18	3	14,667	0,066
CLUSTER20	3	2,818	0,945
CLUSTER12	2	12	0,151
CLUSTER26	2	62,148	0
CLUSTER15	2	13,333	0,101

Tabla 40 Frecuencia de las agrupaciones de palabras en la cuarta contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco

	7. DCc_JRS_PLF3	Chi2	P (2- colas)
CLUSTER09	68	44,909	0
CLUSTER03	26	283,034	0
CLUSTER02	19	178,466	0
CLUSTER05	17	77,745	0
CLUSTER11	17	10,484	0,233
CLUSTER07	16	40,73	0
CLUSTER08	14	27,348	0,001
CLUSTER12	14	12	0,151
CLUSTER04	13	16,246	0,039
CLUSTER06	12	16,71	0,033
CLUSTER14	11	15,677	0,047
CLUSTER13	10	14,382	0,072
CLUSTER10	7	61,202	0
CLUSTER27	7	111,795	0
CLUSTER21	6	65	0
CLUSTER26	5	62,148	0
CLUSTER01	4	21,091	0,007
CLUSTER23	4	146,745	0
CLUSTER16	3	16,286	0,038
CLUSTER17	2	78,387	0
CLUSTER20	2	2,818	0,945
CLUSTER15	2	13,333	0,101
CLUSTER19	1	50,153	0
CLUSTER22	1	47,882	0
CLUSTER24	1	33,333	0

Tabla 41 Frecuencia de las agrupaciones de palabras en la quinta contrarréplica de José Rafael Sañudo, este artículo está firmado con el seudónimo Patricio Leal y Franco

	8. DCc_JRS_PLF4	Chi2	P (2- colas)
CLUSTER03	113	283,034	0
CLUSTER02	78	178,466	0
CLUSTER05	45	77,745	0
CLUSTER09	25	44,909	0
CLUSTER10	15	61,202	0
CLUSTER17	15	78,387	0
CLUSTER04	12	16,246	0,039
CLUSTER16	12	16,286	0,038
CLUSTER11	11	10,484	0,233
CLUSTER07	11	40,73	0
CLUSTER08	11	27,348	0,001
CLUSTER13	9	14,382	0,072
CLUSTER01	9	21,091	0,007
CLUSTER06	8	16,71	0,033
CLUSTER15	8	13,333	0,101
CLUSTER12	7	12	0,151
CLUSTER14	5	15,677	0,047
CLUSTER27	5	111,795	0
CLUSTER18	3	14,667	0,066
CLUSTER21	2	65	0
CLUSTER26	1	62,148	0
CLUSTER23	1	146,745	0
CLUSTER20	1	2,818	0,945
CLUSTER19	1	50,153	0
CLUSTER22	1	47,882	0
CLUSTER24	1	33,333	0

Estos escritos a los cuales se les ha identificado cada grupo, son los que contribuyen al vocabulario diferencial y característico de cada clase descrito anteriormente.

Análisis de la tabla agregada

Se analizó la tabla léxica agregada (vocabulario * grupos de discursos), para agrupar las formas según su distribución.

Las unidades lexicales "típicas" son definidas por exceso (sobre utilización) según un criterio estadístico que requiere la aplicación del Test de Chi cuadrado. Así puede observarse en la Tabla 42, las formas típicas definidas por valores altos del CHI2 y que conforman la "especificidad" del discurso (considerando valores de CHI2>20 en cada uno, se ven las formas más específicas).

Tabla 42 presenta los cluster típicos con sus respectivas palabras de Sergio Elías Ortiz en el primer artículo “HISTORIA?” con el seudónimo “Un indio patriota”

CLUSTER	Palabras	1. DCr_SEO_IP1	Chi2	P (2-colas)
CLUSTER01	alguna, sea, primero, madre			
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	74	44,909	0
CLUSTER21	et, les	23	65	0
CLUSTER15	cada, señores	23	62,148	0
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	13	111,795	0
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	9	61,202	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	8	283,034	0
CLUSTER20	jamás	7	146,745	0
CLUSTER23	jefes, pecados, pueblos, moral	4	194,8	0

CLUSTER17	España, independencia, Julio, Manuel, español, españoles	4	50,153	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	3	178,466	0
CLUSTER06	Colombia, tratado, departamento, razón, he	3	40,73	0
CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	3	27,348	0,001
CLUSTER13	cosa, tanto, dicho, Dios	2	78,387	0
CLUSTER24	humana	2	47,882	0
CLUSTER22	Derecho, gobierno, todas	2	21,091	0,007
CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	1	77,745	0
CLUSTER27	hombre, hombres, humanidad, individuos, sociedad	1	33,333	0

Tabla 43. Presenta los cluster típicos con sus correspondientes palabras, de Sergio Elías Ortiz en el segundo artículo “AÚN VIVE LA HISTORIA” con el seudónimo “Alexo Corfú”

CLUSTER	Palabras	2. DCr_SEO_AC	Chi2	P (2- colas)
CLUSTER01	alguna, sea, primero, madre			
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	36	44,909	0
CLUSTER13	cosa, tanto, dicho, Dios	27	78,387	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	20	283,034	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenía, pesos, puente, entonces, hizo.	10	77,745	0
CLUSTER24	humana	7	47,882	0
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	6	111,795	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	6	178,466	0

CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	5	61,202	0
CLUSTER20	jamás	4	146,745	0
CLUSTER06	Colombia, tratado, departamento, razón, he	3	40,73	0
CLUSTER17	España, independencia, Julio, Manuel, español, españoles	2	50,153	0
CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	2	27,348	0,001
CLUSTER22	Derecho, gobierno, todas	2	21,091	0,007
CLUSTER15	cada, señores	1	62,148	0
CLUSTER23	jefes, pecados, pueblos, moral	1	194,8	0

Tabla 44. Presenta las agrupaciones de palabras típicas de Sergio Elías Ortiz en el tercer artículo “*HISTORIA?*” con el seudónimo “Un indio patriota”

CLUSTER	Palabras	3. DCr_SEO_IP2	Chi2	P (2- colas)
CLUSTER01	alguna, sea, primero, madre			
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	66	44,909	0
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	41	111,795	0
CLUSTER20	jamás	36	146,745	0
CLUSTER23	jefes, pecados, pueblos, moral	34	194,8	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	23	283,034	0
CLUSTER06	Colombia, tratado, departamento, razón, he	21	40,73	0
CLUSTER24	humana	16	47,882	0
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	15	61,202	0

CLUSTER17	España, independencia, Julio, Manuel, español, españoles	13	50,153	0
CLUSTER15	cada, señores	9	62,148	0
CLUSTER27	hombre, hombres, humanidad, individuos, sociedad	8	33,333	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	7	77,745	0
CLUSTER22	Derecho, gobierno, todas	5	21,091	0,007
CLUSTER04	hubo, pueblo, muchos, Nariño.	3	178,466	0
CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	2	27,348	0,001

Tabla 45. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “Los refutadores del señor José Rafael Sañudo” con el seudónimo “Próspero Gallo”

CLUSTER	Palabras	4. DCc_JRS_PG	Chi2	P (2- colas)
CLUSTER01	alguna, sea, primero, madre			
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	56	44,909	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	44	178,466	0
CLUSTER06	Colombia, tratado, departamento, razón, he	25	40,73	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	18	283,034	0
CLUSTER15	cada, señores	17	62,148	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	14	77,745	0
CLUSTER17	España, independencia, Julio, Manuel, español, españoles	9	50,153	0

CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	8	27,348	0,001
CLUSTER21	et, les	8	65	0
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	7	61,202	0
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	6	111,795	0
CLUSTER22	Derecho, gobierno, todas	4	21,091	0,007
CLUSTER24	humana	1	47,882	0
CLUSTER27	hombre, hombres, humanidad, individuos, sociedad	1	33,333	0
CLUSTER13	cosa, tanto, dicho, Dios	1	78,387	0
CLUSTER20	jamás		146,745	0
CLUSTER23	jefes, pecados, pueblos, moral		194,8	0

Tabla 46. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “UNAS VERDADES (1)” con el seudónimo “Patricio Leal y Franco”

CLUSTER	Palabras	5.	Chi2	P (2-colas)
CLUSTER01	alguna, sea, primero, madre	DCc_JRS_PLF1		
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	74	44,909	0
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	39	61,202	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	32	283,034	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	27	77,745	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	21	178,466	0
CLUSTER06	Colombia, tratado, departamento, razón, he	18	40,73	0
CLUSTER22	Derecho, gobierno, todas	14	21,091	0,007

CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	12	27,348	0,001
CLUSTER13	cosa, tanto, dicho, Dios	10	78,387	0
CLUSTER17	España, independencia, Julio, Manuel, español, españoles	7	50,153	0
CLUSTER21	et, les	6	65	0
CLUSTER24	humana	6	47,882	0
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	5	111,795	0
CLUSTER15	cada, señores	3	62,148	0
CLUSTER20	jamás	3	146,745	0
CLUSTER23	jefes, pecados, pueblos, moral	1	194,8	0

Tabla 47. Presenta las agrupaciones típicas palabras de José Rafael Sañudo en el artículo “UNAS VERDADES (2)” con el seudónimo “Patricio Leal y Franco”

CLUSTER	palabras	6.	Chi2	P (2- colas)
CLUSTER01	alguna, sea, primero, madre	DCc_JRS_PLF2		
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	108	283,034	0
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	41	44,909	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenía, pesos, puente, entonces, hizo.	32	77,745	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	32	178,466	0
CLUSTER06	Colombia, tratado, departamento, razón, he	29	40,73	0
CLUSTER17	España, independencia, Julio, Manuel, español, españoles	22	50,153	0
CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	17	27,348	0,001

CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	12	61,202	0
CLUSTER13	cosa, tanto, dicho, Dios	5	78,387	0
CLUSTER22	Derecho, gobierno, todas	4	21,091	0,007
CLUSTER21	et, les	3	65	0
CLUSTER15	cada, señores	2	62,148	0

Tabla 47 Presenta las agrupaciones típicas palabras de José Rafael Sañudo en el artículo “UNAS VERDADES (3)” con el seudónimo “Patricio Leal y Franco”

CLUSTER	Palabras	7.	DCc_JRS_PLF3	Chi2	P (2-colas)
CLUSTER01	alguna, sea, primero, madre	7.			
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	68		44,909	0
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	26		283,034	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	19		178,466	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenía, pesos, puente, entonces, hizo.	17		77,745	0
CLUSTER06	Colombia, tratado, departamento, razón, he	16		40,73	0
CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	14		27,348	0,001
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	7		61,202	0

	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	7	111,795	0
CLUSTER09				
CLUSTER21	et, les	6	65	0
CLUSTER15	cada, señores	5	62,148	0
	Derecho, gobierno, todas	4	21,091	0,007
CLUSTER22				
CLUSTER20	jamás	4	146,745	0
CLUSTER13	cosa, tanto, dicho, Dios	2	78,387	0
	España, independencia, Julio, Manuel, español, españoles	1	50,153	0
CLUSTER17				
CLUSTER24	humana	1	47,882	0
	hombre, hombres, humanidad, individuos, sociedad	1	33,333	0
CLUSTER27				

Tabla 48. Presenta las agrupaciones típicas de palabras de José Rafael Sañudo en el artículo “UNAS VERDADES (4)” con el seudónimo “Patricio Leal y Franco”

CLUSTER	Palabras	8.	Chi2	P (2-colas)
CLUSTER01	alguna, sea, primero, madre	DCc_JRS_PLF4		
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	113	283,034	0
CLUSTER04	hubo, pueblo, muchos, Nariño.	78	178,466	0
CLUSTER05	apuntes, ciudad, siendo, dijo, tenía, pesos, puente, entonces, hizo.	45	77,745	0
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	25	44,909	0
CLUSTER07	debe, haber, varios, hijos, padre, error, primer, tuvo, indio	15	61,202	0
CLUSTER13	cosa, tanto, dicho, Dios	15	78,387	0
CLUSTER06	Colombia, tratado, departamento, razón, he	11	40,73	0

CLUSTER10	José, tienen, ni, Santander, siquiera, Patria	11	27,348	0,001
CLUSTER22	Derecho, gobierno, todas	9	21,091	0,007
CLUSTER09	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	5	111,795	0
CLUSTER21	et, les	2	65	0
CLUSTER15	cada, señores	1	62,148	0
CLUSTER20	jamás	1	146,745	0
CLUSTER17	España, independencia, Julio, Manuel, español, españoles	1	50,153	0
CLUSTER24	humana	1	47,882	0
CLUSTER27	hombre, hombres, humanidad, individuos, sociedad	1	33,333	0

Estudio de la tabla léxica agregada – tipologías de vocabulario según el carácter de artículos

Sobre la tabla léxica agregada con 100 formas en filas y los artículos en columna, se realizó el análisis factorial [AFCS], en este caso todas las filas y columnas forman parte de la nube activa. Se proyecta de manera ilustrativa los tipos de discurso, que según el carácter están clasificados como réplica y contrarréplica.

Tabla 49. Presenta las agrupaciones de palabras según el carácter del artículo

	RÉPLICA	CONTRARRÉPLICA	Chi2	P (2- colas)
CLUSTER01	9	35	5,455	0,065
CLUSTER02	12	194	88,18	0
CLUSTER03	51	297	77,49	0
CLUSTER04	10	51	11,6	0,003
CLUSTER05	18	135	43,24	0
CLUSTER06	9	53	13,97	0,001
CLUSTER07	27	99	13,89	0,001
CLUSTER08	7	62	22,03	0
CLUSTER09	176	264	1,173	0,556
CLUSTER10	29	80	5,52	0,063
CLUSTER11	22	71	7,605	0,022
CLUSTER12	20	40	0,444	0,801
CLUSTER13	21	34	0,011	0,995
CLUSTER14	17	45	2,688	0,261
CLUSTER15	5	19	2,844	0,241
CLUSTER16	24	32	0,686	0,71
CLUSTER17	29	33	2,275	0,321
CLUSTER18	16	8	8,711	0,013
CLUSTER19	19	40	0,706	0,703
CLUSTER20	3	8	0,491	0,782
CLUSTER21	23	25	2,222	0,329
CLUSTER22	25	9	18,83	0
CLUSTER23	47	8	53,97	0
CLUSTER24	9	3	7,2	0,027
CLUSTER25	39	1	61,44	0
CLUSTER26	33	28	7,17	0,028
CLUSTER27	60	23	42,86	0

Con base en la tabla 50 se elaboró la tabla 51 para establecer un listado ordenado con las agrupaciones lexicales (cluster) típicas según el carácter de los artículos.

Las agrupaciones lexicales (cluster) típicas son definidas por exceso (sobre utilización) según un criterio estadístico que requiere la aplicación del Test de Chi cuadrado. Así puede observarse en la Tabla 4.50, las formas típicas definidas por valores altos del CHI2 y que conforman la “especificidad” del discurso (considerando valores de $CHI2 > 20$ en cada uno, se ven las formas más específicas).

Tabla 50 Asociaciones de palabras típicas según carácter de los artículos

	RÉPLICA	CONTRARRÉPLICA	Chi2	P (2- colas)
CLUSTER02	12	194	88,183	0
CLUSTER03	51	297	77,49	0
CLUSTER25	39	1	61,44	0
CLUSTER23	47	8	53,965	0
CLUSTER05	18	135	43,235	0
CLUSTER27	60	23	42,86	0
CLUSTER08	7	62	22,03	0

En la réplica aparecen en orden de frecuencia las siguientes agrupaciones típicas de palabras

Tabla 51 Agrupaciones típicas de palabras en los artículos de Sergio Elías Ortiz

	RÉPLICA	Chi2	P (2- colas)
CLUSTER27 hombre, hombres, humanidad, individuos, sociedad	60	42,86	0
CLUSTER03 cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	51	77,49	0
CLUSTER23 jefes, pecados, pueblos, moral	47	53,965	0
CLUSTER25 responsabilidad, solidaridad, sistema	39	61,44	0

CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	18	43,235	0
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	12	88,183	0
CLUSTER08	Bolívar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	7	22,03	0

Por su parte estas son las agrupaciones típicas de palabras utilizadas por José Rafael Sañudo en sus contrarréplicas

Tabla 52 Agrupaciones típicas de palabras de José Rafael Sañudo en su contrarréplica

	CONTRARRÉPLICA	Chi2	P (2- colas)	
CLUSTER03	cabildo, capitán don, Juan, gobernador, Popayán, Pasto, era, sino, misma, nombre, Francisco, san, obispo, Quito, lugar, siguiente	297	77,49	0
CLUSTER02	año. Boletín, número, página, dice, Ortiz, señor, conjetura, da	194	88,183	0

CLUSTER05	apuntes, ciudad, siendo, dijo, tenia, pesos, puente, entonces, hizo.	135	43,235	0
CLUSTER08	Bolivar, vida, estudios, doctor, Sañudo, historia, hacer, historiador, obra hechos, libertador	62	22,03	0
CLUSTER27	hombre, hombres, humanidad, individuos, sociedad	23	42,86	0
CLUSTER23	jefes, pecados, pueblos, moral	8	53,965	0
CLUSTER25	responsabilidad, solidaridad, sistema	1	61,44	0

Esto permitió:

1 Obtener la matriz de factores del vocabulario que representa la estructura principal de la información contenida en la tabla léxica, las relaciones entre formas y su distribución según el carácter de los artículos.

2 Aplicar sobre esta matriz el procedimiento de clasificación jerárquica directa a través del método de Ward

3 Seleccionar cluster (clases óptimas según la relación de inercia intra dentro del grupo e inercia intra entre clases) y optimizar la partición mediante el algoritmo k- mean.

Para una mejor interpretación nos ayudamos con los procedimientos de especificidades para describir y validar el vocabulario típico de cada clase.

Descripción y validación estadística del vocabulario característico de cada clase

Se identificó las características “típicas” de cada cluster, con una validación superior al 95% de certeza ($\alpha=0.05$), para Valores de Test mayores o iguales a 1.98; y un nivel de significación del 99% ($\alpha=0.01$), para Valores de Test mayores o iguales a 2.58.

ESBOZO DE INFORME FINAL CON CONCLUSIONES

Edición y partición del Corpus

La estrategia de describir para cada artículo el carácter, autor, seudónimo, permitió incluir esta diferenciación en el identificador del glosario de formas gráficas.

La primera información que se obtuvo sobre el corpus, fue la repartición de las palabras, además su longitud y el tamaño. Se obtuvieron 5402 formas simples, de un total de 21646 ocurrencias; 1858 formas con al menos 2 frecuencias, que corresponden a 18103 ocurrencias (88.44% del corpus), 3543 formas de frecuencia 1 (harpas) (que corresponde a 11.56% del corpus sin lematizar) Entre las palabras claves del dominio (con significado), las más frecuentes son: “Bolívar”, “Sañudo”, “doctor”, “historia”, “estudios”, “vida”, “Pasto”.

El contexto obtenido mediante el estudio de concordancias de las formas más típicas y características de los discursos, permitió encontrar información cognitiva transmitida por los escritores. En el estudio de concordancias se encontraron algunas definiciones, entre la más importante, la diferencia según el contexto de la palabra “justicia”, en algunos casos hace referencia al atributo del soberano (Bolivar elevado al rango de tal) y en otros se alude al sentimiento de rechazo frente a aquello que se considera atenta al respeto de lo instituido, o a la venganza por la profanación.

Habiéndose quitado las palabras herramientas (artículos, conjunciones, preposiciones y pronombres) que no tienen significado en el Dominio del Corpus, se obtuvo el glosario de palabras y segmentos con 4826 formas, que es el que se analiza en primer término.

El Estudio de la Tabla Léxica Base

Observando la representación gráfica, en el plano de los individuos representados por los discursos, se revela una distribución que generalmente varía en las dos dirección del eje horizontal (1º factor) y vertical (2º factor), debido a que algunos discursos contribuyen más que otros en el vocabulario del Corpus, presentando más relevancia.

Es notorio que existen grupos de discursos con un vocabulario semejante (homogeneidad dentro del grupo) y una buena heterogeneidad entre grupos (diferente vocabulario).

Una interpretación de la distribución del vocabulario en los discursos se ve en el plano de los discursos y vocabulario más relevante asociado. Como hecho, puede observarse globalmente en representación cuatro grupos destacados de discursos distribuidos según el vocabulario común de cada uno:

1. Un grupo de enunciados a la izquierda del gráfico, que se asocia a: “respeto”, “sueños”, “testimonios”, “esperanzas”, “trabajo”, “dignidad”, “hombre”, “vida”, indicando mensajes de ilusión y ánimo.
2. En forma muy diferenciada, otro grupo de enunciados a la derecha arriba del gráfico, indica otra tendencia en los mensajes, de reproche al pasado histórico de: “justicia”, “memoria”, “democrático”, “histórico”, “quedar”, “impunidad”, “verdad”.
3. Al centro se encuentran los mensajes con alto contenido de vocabulario que evoca el conflicto: “autoridades”, “honor”, “profundos”, “olvidar”.
4. Abajo a la derecha, se ven dos grupos de discursos contiguos, uno revela mensajes referidos al futuro del perpetrador y el otro grupo de discursos en el extremo inferior, que reflexiona acerca de la experiencia “solidaridad” “humanidad”

La Clasificación realizada por el método de Ward y su consolidación por el criterio de las k-mean, considerando a los artículos y su vocabulario como elementos activos y los datos complementarios asociados como ilustrativos, ha permitido ahora encontrar grupos diferenciados muy homogéneos, de clases óptimas según la relación de inercia intra, dentro del grupo y la inercia inter, entre clases y su optimización. Así pueden ser explicadas las tipologías encontradas y validadas estadísticamente por pruebas del Valor Test.

En una primera aproximación se puede interpretar que aunque se esperaba que las palabras típicas en las réplicas de Sergio Elías Ortiz, estuviesen relacionadas con la impugnación a la concepción de historia de Sañudo (el problema de la expiación) sin embargo el de mayor frecuencia es “solidaridad”, como si la inquietud fuese de carácter moral por la existencia o no de solidaridad en los pecados de los líderes y la responsabilidad de los ciudadanos frente a esto; entonces las inquietudes están enmarcadas en las consecuencias “peligrosas” de aceptar como ciertos los presupuestos de Sañudo. Por su parte la forma “señor” típica de Sañudo, respondería a la necesidad de plantear el debate en condiciones de urbanidad con un trato formal, puesto que como consecuencia de sus Estudios, la reacción frente a él, ha implicado la declaración como antipatriota y el peligro de ser atacado desconociéndole su derecho a libertad de opinión y conciencia.

Evidentemente la progresión del vocabulario se ve asociado a factores como la fecha y secuencia de la disertación que presenta el artículo.

Dos palabras empleadas más frecuentemente en los mismos discursos tendrán una posición próxima en el gráfico. Dos discursos con vocabulario similar serán próximos en el gráfico.

La posición de un discurso a todas las palabras se da por las distancias existentes al mismo

Este fenómeno ratifica la existencia de una variación progresiva del vocabulario según el motivo y fecha del artículo.

La distribución del vocabulario y su diferenciación en los artículos analizados, alcanzan finalmente una buena significación, descubriéndose “patrones” o ideas claves que sintetizan la comunicación planeada por el disertante, que termina preocupándose por las consecuencias sociales y humanas por la caída del héroe. Al tener que dar cuenta de los aspectos relevantes obtenidos y de difícil formalización, como son todos los aspectos relacionados con el ámbito de la pragmática (deseos, creencia, opiniones, intenciones, ideas cognitivas), el análisis de los textos, obtenido, presenta ventajas insustituibles a cualquier otro método, permite en el análisis y crítica salir de la condición de ser juez y parte en el contexto de la polémica y encontrar el núcleo del conflicto.

Es inquietante la relación entre “poder soberano” (dato resultante dentro del semema “justicia”) y muerte (a la cual está predestinado el perpetrador). Foucault nos ayudaría a revisar esta relación; en El capítulo intitulado Derecho de muerte y poder sobre la vida, el cual hace parte del primer tomo de la historia de la sexualidad más conocido como la voluntad de saber, en él se propone su tesis del paso del derecho de muerte al poder sobre la vida, como la gran transformación del Estado moderno: “La vieja potencia de la muerte, en la cual se simbolizaba el poder soberano, se halla ahora cuidadosamente recubierta por la administración de los cuerpos y la gestión calculadora de la vida”

La potencia de la muerte en la cual se simbolizaba el poder soberano estaba influenciada por la patria potestas romana, según la cual el padre podía disponer a voluntad de la vida y/o de la muerte de sus hijos y esclavos. Este mismo derecho parece ser heredado, guardadas proporciones, por el soberano respecto de sus súbditos, este derecho a disponer de la vida y de la muerte de los súbditos se presenta como “un derecho “disimétrico”, pues, no se ejerce, sobre la vida de la misma manera que sobre la muerte; o, mejor dicho, sólo se ejerce sobre la vida mediante la amenaza más o menos inminente de su aniquilación, esto es, a través de la muerte: es el derecho de hacer morir o de dejar vivir. Además, este “derecho de vida y de muerte”, fundamento último de la soberanía, era el resorte extremo de toda una serie de mecanismos de poder esencialmente basados en la

deducción, la sustracción, la apropiación (de los bienes, de los servicios, de los impuestos, del tiempo y, finalmente de las vidas) de los súbditos”

La lógica del poder soberano legitima “la autoridad del gobernante como una instancia cuyo fin fundamental y cuya tarea es alcanzar el bien común y la salvación de todos”

Es en la teoría clásica de la soberanía, donde aparece el derecho de vida y muerte como uno de los atributos fundamentales del soberano. Sin embargo, este derecho sobre la vida y la muerte ha de ser precisado desde la teoría, pues, preguntarnos ¿Qué significa en los hechos derecho de vida y de muerte? equivale a decir que el soberano tiene la potestad de hacer morir o dejar vivir según su voluntad. Esto implica, según que la vida y la muerte no son tomados como fenómenos naturales, inmediatos, de algún modo originarios o radicales, parecen extraños al campo del poder político. Lo que llevado hasta los límites del análisis implicaría admitir que en las confrontaciones del poder, el sujeto no es sujeto pleno de derecho ni vivo ni muerto. Estaríamos entonces haciendo referencia a un sujeto neutro, que sólo gracias a la voluntad del soberano tiene el derecho de estar vivo o de estar muerto. En cualquier caso, la vida y la muerte de los sujetos pasan a convertirse en derechos como consecuencia de una decisión emanada del soberano.

Este derecho de vida y de muerte, como anotamos antes, no se equipara de manera literal con la patria potestas romana, pues, este llamémosle “privilegio” otorgado al soberano para disponer de la vida de sus súbditos, ha de ser respaldado por una situación excepcional donde el soberano, y por tanto, el bienestar colectivo vean seriamente comprometida su estabilidad. Una vez que el soberano pide a sus súbditos que tomen parte en la defensa del Estado podemos decir que “es lícito para él exponer sus vidas: en ese sentido ejerce sobre ellos un derecho indirecto de vida y de muerte”

Si es uno de sus súbditos el que se levanta contra él, entonces el soberano puede ejercer sobre su vida un poder directo: a título de castigo lo matará. Es decir, el soberano podrá decidir entre hacer morir o dejar vivir a quien se levanta u opone al poder que le ha sido otorgado por sus súbditos.

De cualquier manera en que el soberano ejerza el derecho de vida y de muerte, ya sea en la forma absoluta venida de la patria potestas romana o con las restricciones modernas, como los derechos de los individuos frente a sus gobernantes, estaremos en presencia de un derecho disimétrico dado que: “El soberano no ejerce su derecho sobre la vida sino poniendo en acción su derecho de matar, o reteniéndolo; no indica su poder sobre la vida sino en virtud de la muerte que puede exigir. El derecho que se formula como de vida y de muerte es en realidad el derecho de hacer morir o de dejar vivir. Después de todo, era simbolizado por la espada. Siempre que se halle amenazada la existencia del

soberano, éste tendrá el derecho de hacer uso de esas existencias en función del interés mayor que es conservar su propia autoridad, así como la lealtad y la sumisión de sus súbditos. Este derecho de vida y de muerte, venido de la espada, supondrá siempre un ejercicio desde la muerte puesto que “el súbdito ante el poder soberano no se encuentra, por pleno derecho, ni vivo ni muerto, su estado es completamente neutro ante una instancia que se define por la posibilidad de matarlo o no”

Al respecto se resaltaría la transformación más significativa del poder soberano, el desplazamiento del derecho que tiene el soberano a defenderse o exigir ser defendido, hacia el derecho del cuerpo social a asegurar su vida, mantenerla y desarrollarla. Así, el antiguo poder soberano de matar y exigir sumisión y lealtad al soberano aparece enmascarado y, a su vez, unido a las estrategias del poder (en este caso académico) como complemento de un nuevo poder que aspira a la administración de la vida: el biopoder.

La forma disciplinaria del poder es una forma de poder que, básicamente, no actúa sobre los sujetos, sino que los encausa hacia un horizonte de acción: No disciplina, sino que normaliza. No funciona con base en los ordenamientos jurídicos o los derechos, sino en normas y estándares que aluden a una tecnología social. Es un poder que surge con el desarrollo de las ciencias humanas y, en particular, con el de las ciencias de la normalización. De tal modo que el poder no está centrado, sino que es difuso, no es propiedad exclusiva de nadie, sino anónimo, no se ejerce sino que se trasmite y se vive.

Una tecnología política implica el despliegue de determinadas técnicas de intervención con el propósito de modificar la vida, tanto de individuos como de poblaciones. Es decir, es una forma aplicada de conocimientos, saberes y prácticas de intervención, dirigidas a cambiar la conducta de los individuos con el objetivo de tener un control regulado de los mismos al interior de la sociedad.

La tecnología política tiene el objetivo de corregir la “anomalía” de los individuos y ajustar sus campos de acción a lo socialmente deseable y adecuado.

La tecnología del yo hace relación a las prácticas por las cuales los individuos se constituyen así mismo como sujetos, dentro de un sistema atravesado por un sinnúmero de relaciones de poder. Las tecnologías del yo entran en acción cuando, en la aplicación de una tecnología política, el individuo es conducido a tomar conciencia de su “anormalidad” o “falta” pero es inconsciente de aquello a lo que se induce. Así los sujetos solicitan, por iniciativa propia, ser tratados nuevamente por una tecnología política, con sus respectivas técnicas de intervención.

La gubernamentalidad se expresa en medidas estatales que operan sobre la producción de conocimientos, cuya elaboración emana de un conjunto de

instituciones, expertos y disciplinas destinados a constituir un régimen social normalizador. En ese sentido, todo ese aparato de saberes e instituciones sociales va dirigido a identificar, clasificar y tipificar el grado de “anormalidad” de los individuos con el propósito de normalizarlos, modificarlos y acomodarlos a un orden social concebido por el discurso emanado del poder del Estado

Estas medidas gubernamentales no tienen el propósito último de castigar o reprimir, sino de administrar la conducta de los individuos actuando sobre sus posibilidades de acción, “concientizándolos” de tales posibilidades y anormalidades.

5. CONCLUSIONES

El estudio del discurso, concebido desde una perspectiva teórico-metodológica específica, nos permite conocer y describir no solamente lo que dice el emisor de determinados discursos sino, también, el contexto y la situación coyuntural en que éstos son emitidos. El discurso no nos proporciona por sí solo toda la información necesaria para conocer la realidad social, pero sí nos permite encontrar claves que nos llevan a la reconstrucción de esa realidad.

Como señala Bourdieu, "El trabajo político se reduce, en lo esencial, a un trabajo sobre las palabras, porque las palabras contribuyen a construir el mundo social" (En Eribon 1982).

Es importante la inclusión del estudio del discurso como una propuesta metodológica ya que su uso para examinar, por ejemplo, los discursos sobre lo público y lo político, ofrece una perspectiva que tiene que ver más con la tarea de descubrir lo que es importante para la sociedad en términos de **valores** en lugar de políticas, y de **visiones o representaciones**, en lugar de programas. Desde esta perspectiva, se trata de redimensionar la manera en que el lenguaje actúa - con una eficacia particular- en la vida social y en la historia de los seres humanos.

Desde un punto de vista se asume que el lenguaje es un portador de contenido político y no solamente una herramienta para poder hablar sobre fenómenos extradiscursivos que existen independientemente de lo que decimos y, por lo tanto, de lo que concebimos y afirmamos. De ahí que, desde esta perspectiva, el estudio del discurso político implique en parte, y en cierto sentido, estudiar el lenguaje en el mundo social, la manera en que el lenguaje es usado en la vida social cotidiana y los modos en que los múltiples y variados usos del lenguaje se entrecruzan con el poder, alimentándolo, sosteniéndolo y ejecutándolo. En otras palabras, se busca poner en evidencia las maneras en que ciertas relaciones de poder son mantenidas y reproducidas en un conjunto interminable de expresiones que movilizan el sentido en el mundo social. Por ello, es necesario reconocer que aunque la ideología se manifiesta de muchas formas (por ciertas prácticas sociales, por ciertas instituciones, por símbolos, etc.), su dominio privilegiado, el lugar donde ejerce directamente su función, es el lenguaje.

Existen diferentes propuestas teórico-metodológicas para el análisis del discurso que remiten a diferentes paradigmas y que implican diferentes procedimientos de análisis. No es el propósito de este trabajo dar un recuento de esas diferentes propuestas sino exponer el enfoque que se ha adoptado y que se ha puesto en práctica.

Los métodos análisis comparativo de Texto, proporcionan herramientas extraordinarias para poder extraer la información contenida en textos. Es el

procedimiento de análisis que más se aproxima a la realidad. Como siempre ocurre, esta metodología no está exenta de dificultades. Cuando se trata de comprimir miles de palabras en unos resultados concisos, siempre hay una simplificación que puede producir alguna deformación. Por otra parte, como manifiesta L. Lebart, cada análisis textual es una verdadera investigación.

Aunque se ha expuesto una guía metodológica de análisis, ésta no es totalmente automática, el investigador dispone de muchas opciones y tiene que tomar decisiones no excluyentes o realizar el análisis de varias formas diferentes para comparar los resultados. Aquí el arte y la experiencia del investigador enriquecen el estudio.

El "purismo" metodológico de esta propuesta suele despertar una gran suspicacia entre quienes piensan que un objeto de lenguaje no puede ser tratado como una simple matriz de datos. Si bien esta crítica es, en una gran medida, legítima, no obsta que el Análisis Estadístico de Datos Textuales ha abierto vías extremadamente interesantes y estimulantes en lo que hace al tratamiento del corpus (y a la elaboración de protocolos de descripción y de clasificación de objetos textuales).

El desarrollo de las técnicas de la estadística textual ha hecho que el análisis estadístico de textos se haya constituido en una herramienta interdisciplinar, integrada por: la estadística, el análisis del discurso, la lingüística, la informática, la investigación documental; y es cada vez más utilizada en diversos campos de las ciencias sociales: historia, política, economía, sociología, psicología, etc. Y específicamente en el análisis de los discursos sociales en la investigación del consumidor, del ciudadano, y en general, del sujeto mediático.

Las técnicas de análisis de datos desarrolladas a partir de las aportaciones de Jean Paúl Benzécri han permitido el análisis de matrices de datos, aplicación del Análisis Factorial a tablas de contingencia de n (filas) \times p (columnas) a partir grandes matrices de datos extendidas y la visualización de los resultados en un mapa perceptual. Se trata ahora de hacer que el análisis de datos devenga efectivamente un proceso de adquisición y de apropiación de los conocimientos expresados bajo la forma de datos textuales.

El análisis textual puede ayudarnos a establecer hipótesis de trabajo iniciales que faciliten la identificación de patrones, la iluminación de relaciones entre conceptos o variables, en suma, a enriquecer la comprensión sobre un fenómeno objeto de estudio. Efectivamente, con el análisis textual podemos establecer hipótesis novedosas o rivales, explorar nuevas relaciones o perspectivas sobre un problema

No obstante, también puede utilizar el análisis textual en la comprobación de hipótesis previas. Es decir, además de favorecer la formulación de hipótesis, el análisis textual puede confirmar hipótesis o supuestos de partida cuando existe

una teoría previa. Desde esta perspectiva, la identificación de los segmentos de texto puede hacerse en función de un referente teórico, que también puede ser la referencia fundamental en la interpretación de los ejes factoriales y las clases.

La aproximación estadística a los datos textuales presentada en este trabajo ofrece una nueva lectura de los textos, lectura esencialmente distinta pero complementaria de la lectura humana. Dicha lectura proporciona una descripción cuantitativa, sistemática y exhaustiva del vocabulario.

5.1. MEJORÍAS

La guía para el enfoque de estudio comparativo de textos, ofrece una aproximación comparativa: se describen, analizan e interpretan las diferencias entre los textos. Se basa en los principios de la estadística textual y constituye un área de trabajo y de investigación en pleno desarrollo, situado en la encrucijada de los ámbitos de investigación relacionados con el estudio de textos. Aporta herramienta de naturaleza estadística y, como tal, ofrece un enfoque comparativo destinado a poner de relieve las diferencias entre los textos estudiados. Opera a partir del recuento exhaustivo de las palabras y los segmentos, pero teniendo en cuenta que los mismos operan dentro de un universo que representa el documento analizado.

La transparencia del contenido respecto al análisis permite descifrar la información sobre el significado de las palabras, según los objetos observados, contenida en el propio documento. En el momento de interpretar los resultados se reintroduce lo que se conoce de la lengua, de las palabras y de los propios textos. Ello permite transformar lo cuantificable en significativo y volver al contenido al final del análisis.

Por lo tanto, la riqueza de esta propuesta de análisis textual se encuentra en que ofrece la posibilidad de explorar el “texto bruto”, evitando que las preconcepciones del investigador “contaminen” el análisis. Su utilidad para el análisis de todo tipo de materiales discursivos.

Estos métodos de análisis de texto, se enriquecen, al utilizarlos para el estudio de la riqueza léxica y estructural temporal.

Los ejemplos presentados exponen las posibilidades de utilización de las técnicas provenientes del análisis factorial y de la técnica de palabras asociadas, que consiente la obtención de lo que puede identificarse como la dinámica presente entre las temáticas mencionadas en los textos, para algunos tipos de textos especiales.

5.2. RUMBOS DE INVESTIGACIÓN POSTERIOR

En este trabajo se expuso un esquema general para hacer análisis estadístico de texto, aunque los esfuerzos se concentraron en la etapa de descubrimiento. Por ello, gran parte del trabajo futuro que se presenta a continuación considera el desarrollo de las demás etapas del proceso. Pero, también se plantea la necesidad de buscar más y mejores métodos que permitan integrarse en las guías propuestas para facilitar el proceso de explotación de este tipo de datos textuales. El desarrollo en el futuro de la propuesta implica en su enfoque la articulación de tres ejes fundamentales: el discurso, la ideología y el poder. Se señalarán algunas reflexiones en torno a dichos conceptos.

En la perspectiva del estudio del discurso político la producción ideológica se puede dar el lujo de disfrazar, desplazar o desviar los conflictos o la potencialidad de los mismos, puede incluso acrecentarlos o atenuarlos al articular una disputa imaginaria en las potencialidades afectivas (Ansart, 1983)

De ahí que considere a la ideología como un instrumento permanente de los poderes y como el espacio simbólico en el cual éstos se legitiman o impugnan, se refuerzan o debilitan continuamente.

Siguiendo a Eagleton, la ideología tiene que ver con el discurso -con ciertos efectos discursivos concretos. Representa los puntos en que el poder incide en ciertas expresiones y se inscribe tácitamente en ellas. El concepto de ideología pretende revelar algo de la relación entre una expresión y sus condiciones materiales de posibilidad, cuando se consideran dichas condiciones de posibilidad a la luz de ciertas luchas de poder centrales para la reproducción (o también, para algunas teorías, la contestación) de toda una forma de vida social (Eagleton, 1997) Desde esta línea de reflexión, otra cuestión a destacar es la de los modos generales de operación de la ideología y las formas en que éstos se pueden vincular con ciertas estrategias de construcción simbólica. Aun cuando todavía queda por realizarse una investigación, más profunda, sobre la manera en que la ideología opera en ciertas condiciones socio-históricas específicas, ciertas modalidades de su funcionamiento han sido identificadas como un primer avance. Por ejemplo, Thompson (Thompson, 1993) distingue cinco modos de operación de la ideología: la legitimación, la disimulación, la unificación, la fragmentación y la reificación o cosificación (pp. 66-73).

Con respecto a la primera, cabe señalar que las relaciones de poder se mantienen si se apoyan en la **legitimación**. Un sistema de dominación puede ser mantenido, como observa Weber (Weber, 1978), al ser representado como legítimo, es decir, como un sistema que es justo y digno de apoyo. Esta legitimación se logra a través de apelar a fundamentos racionales, tradicionales o carismáticos, los cuales, valdría la pena añadir, se expresan generalmente por medio del lenguaje.

La ideología también puede operar a través de la **disimulación o el encubrimiento**.

Las relaciones de poder que sirven a los intereses de unos a expensas de los demás pueden ser ocultadas, negadas o bloqueadas de varias maneras, por ejemplo, describiendo los procesos o acontecimientos sociales con ciertos términos que pongan de relieve algunos rasgos en detrimento de otros, o al representar o interpretar dichos procesos de una manera que disimula o encubre lo que realmente son.

Una tercera modalidad es la **unificación**. Las relaciones de poder pueden ser establecidas y sostenidas al construir, en el nivel simbólico, una forma de unidad que abarca a todos los individuos de una identidad colectiva, a pesar de las diferencias y divisiones que pueden separarlos. Una estrategia típica de esta modalidad, expresada por medio de formas simbólicas, es la estrategia de la estandarización.

La cuarta modalidad es la **fragmentación**. Las relaciones de poder pueden ser mantenidas movilizándolo de tal forma que fragmente a los grupos y ubique a los individuos y a las facciones en oposición. "Divide y gobierna" es una conocida estrategia de los grupos dominantes, aunque, a menudo, los procesos de fragmentación son menos intencionales de lo que sugiere esta máxima.

Una última modalidad es la **reificación o cosificación**. La ideología puede operar al representar un estado de cosas transitorio e histórico como si fuera permanente, natural y atemporal. El restablecer la dimensión de la sociedad "sin historia", como señala Claude Lefort (Lefort, 1986), es una característica clave de la ideología de las sociedades modernas.

Thompson (Thompson, 1993) también presenta algunas de las maneras en que dichos modos se pueden vincular con diversas estrategias de construcción simbólica. Introducir el concepto de discurso es abrir una ruta para la investigación de la relación entre lenguaje, ideología y poder. Se trata de un concepto que ha sido utilizado ampliamente y del que, en cierta manera, se ha abusado mucho en discusiones recientes, en parte porque se deriva de numerosas y variadas fuentes y debates. Es necesario especificar algunos de los rasgos fundamentales del discurso.

Primeramente, hay que recordar el carácter social del lenguaje y el carácter activo del uso del lenguaje. Austin (Austin, 1962) señaló que producir un enunciado es entablar un cierto tipo de interacción social, y que hablar es una manera de actuar y no simplemente de informar o describir lo que se hace. Pero, además, puntualizó que para la realización de ciertos "actos de habla" es esencial que la persona que los emite tenga el poder (ya sea institucional, social o familiar) para ejecutarlos; es decir, ciertos actos de habla son inseparables de una institución, de aquella que el

acto presupone. Dicha concepción permitió superar el modelo puramente comunicacional y avanzar hacia una concepción más sociológica del discurso. Wittgenstein (Wittgenstein, 1953), por su lado, puso de relieve que las expresiones funcionan sólo en el contexto de juegos de lenguaje en los que participa (y debe participar) más de un individuo y que, por tanto, constituyen en algún sentido formas de la vida social. Estas observaciones han servido de base para los estudios que intentan poner de relieve lo que está en juego si se considera al lenguaje como un fenómeno social, es decir, como un fenómeno inmerso en relaciones de poder, en situaciones de conflicto y en procesos de cambio social.

Existen varias definiciones e interpretaciones del concepto discurso, dependiendo de la posición teórica que uno adopta. Siguiendo a (Pecheux, 1978) y Robin (Robin, 1973), se entiende por discurso "toda práctica enunciativa considerada en función de sus condiciones sociales de producción, que son fundamentalmente condiciones institucionales, ideológico-culturales e histórico coyunturales".

No es fácil caracterizar la especificidad del discurso político. Una primera dificultad es que lo político y lo ideológico, dos de sus rasgos fundamentales, pueden encontrarse en casi todo tipo de discurso. Verón (Verón, 1987), al señalar que existe una serie de dificultades en el intento de describir lo que es el discurso político, sugiere que al abordar la caracterización de un tipo de discurso, se deben trabajar simultáneamente los niveles que a continuación se enumeran:

a) En primer lugar, lo que se trata de conceptualizar no es nunca un discurso, sino un **campo discursivo**. Esto implica que lo que se trata de construir no es una tipología de discursos, sino una tipología de juegos de discurso. Desde un inicio nos vemos confrontados con el análisis de procesos de intercambio discursivo.

b) En segundo lugar, y en consecuencia, la definición de un "tipo" supone la definición de una serie de variantes del mismo, que no son otra cosa que diferentes estrategias dentro del mismo juego.

c) En tercer lugar, la descripción de intercambios discursivos implica que se trabaja en diacronía: los intercambios ocurren en el tiempo y una misma estrategia varía a lo largo del tiempo. Por lo tanto, aun en el plano de la caracterización de una estrategia discursiva, se nos plantea el mismo problema de diferenciar un "núcleo" invariante y un sistema de variaciones.

d) En cuarto lugar, los diferentes modos de manifestación de un cierto "tipo" de discurso no pueden ser dejados de lado: los discursos sociales aparecen materializados en soportes significantes que determinan las condiciones de su circulación: la escritura de la prensa, la oralidad de la radio, etc. Es evidente que no podemos analizar de la misma manera los discursos políticos que aparecen en eso diferentes medios.

La descripción de un “tipo” supone la descripción de múltiples estrategias, de procesos de intercambio, de variaciones de cada estrategia a lo largo de un proceso discursivo, de modificaciones de las estrategias según el soporte significante.

De acuerdo a Verón, es necesario diferenciar, a través de esta maraña de niveles que se interdeterminan, lo esencial de lo accesorio, lo que es específico del discurso político de lo que no lo es, vale decir, los elementos que constituyen el “núcleo” del juego discursivo político, de aquellos elementos que pueden manifestarse en dicho juego pero que aparecen también en otros juegos de discurso que no son el político.

En la actualidad, otra cuestión que debe tomarse en cuenta en relación con la determinación de lo que es un discurso político es la función de los medios de comunicación.

Como se pregunta Bonnafus (Bonnafus, 1998) “¿Existe hoy en día algún discurso político ‘puro’ que no pase por los medios de comunicación?”. Ella señala que en realidad muy pocos, por eso trabajar sobre el discurso político hoy, es casi siempre como trabajar sobre el discurso “filtrado” (en el sentido de Chomsky) por los medios de comunicación y tener en cuenta por lo tanto su lógica comunicacional.

Además, hablar de discurso político supone necesariamente que existen discursos que no son políticos, lo cual presupone ciertas hipótesis sobre una tipología de discursos sociales, tipología que no existe todavía. Sin embargo, podemos reconocer en la bibliografía existente sobre el tema dos tipos de concepciones sobre lo que es el discurso político:

a) Concepción restrictiva (en sentido estricto o institucional). “Es el discurso producido dentro de la ‘escena política’, es decir, dentro de los aparatos donde se desarrolla explícitamente el juego del poder” (Giménez, 1983). Así, son ejemplos del discurso político, en sentido estricto, el discurso presidencial, el de los partidos políticos, el de la prensa política especializada, el discurso emitido por los medios electrónicos en ciertos momentos y, en algunos casos, el magisterial, el del ejército y la policía.

b) Concepción extensiva. La diferencia con la anterior es que se basa en un concepto ampliado de “la política”, que da cabida a aquellos discursos que, si bien no son emitidos desde los lugares institucionales donde se da el juego del poder, tienen una intención política; es decir, tienen como objetivo incidir en las relaciones de poder existentes. En esta concepción, el discurso de la disidencia sería también considerado como discurso político.

Aunque actualmente no existen caracterizaciones del discurso político que tomen en cuenta todos los niveles antes señalados, sí se han determinado algunas características formales del discurso político⁶²:

a) es un discurso que no se dirige tanto a convencer al adversario, como supone la retórica tradicional, sino a reconocer, distinguir y confirmar a los partidarios y atraer a los indecisos;

b) es un discurso estratégico, en la medida en que define propósitos, medios y antagonistas;

c) manifiesta propiedades performativas, lo que significa que quien lo sustenta no se limita a informar o transmitir una convicción, sino que también produce un acto, expresa públicamente un compromiso y asume una posición;

d) tiene una base esencialmente polémica: la enunciación política parece inseparable de la construcción de un adversario;

e) es un discurso argumentado que se presenta como un tejido de tesis, argumentos y pruebas destinados a esquematizar y teatralizar, de un modo determinado, el ser y el deber ser políticos ante un público determinado y en vista de una intervención sobre este público.

⁶² Véase, Giménez (1983) págs.126 y ss.

BIBLIOGRAFIA

Alba, M. (2004). *Papers on Social Representations*. Recuperado el 12 de Agosto de 2010, de Papers on Social Representations: <http://www.psych.lse.ac.uk/psr/>

Ansart, P. (1983). *Ideología, conflictos y poder*. México: Premiá.

Arco, L., Bello, R., Meeros, J. M., & Pérez, Y. (2006). Agrupamiento de Documentos Textuales mediante Métodos Concatenados. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*. No.30, ISSN: 1137-3601. , 43-53.

Armony, V. (2002). *Les Textes De Méthodologie, Chaire de Recherche du Canada en Mondialisation, Citoyenneté et Démocratie*. Recuperado el 3 de Diciembre de 2010, de Textes De Méthodologie, Chaire de Recherche du Canada en Mondialisation, Citoyenneté et Démocratie.: <http://www.chaire-mcd.ugam.ca/>

Austin, J. (1962). *How to do things with words*. Oxford: Clarendon Press.

B, E., & Pagès, J. (1992). *Análisis factoriales simples y múltiples, Objetivos, métodos e interpretación*. Bilbao: Servicio Editorial Universidad del País Vasco.

Bécue, M. (1997). Análisis Estadístico de Textos. *Cuarto seminario de capacitación de docentes, PRESTA*, . Universidad de Concepción de Chile y Universidad libre de Bruxelles, Belgique.

Bécue, M. (1991). *Análisis de Datos Textuales, Métodos estadísticos y algoritmos*. París: CISIA.

Benzécri, J. (1973). *La taxonomie. Voll. L'Analyse des Correspondances, Vol II.* . Paris: Dunod.

Benzécri, J. (1981). Pratique de l'Analyse des Données. En J. y. Benzécri, *Linguistique et Lexicologie. tome III*. Paris: Dunod.

Bolasco, S. (1999). *Analisi Multidimensionale Dei Dati. Metodi, Strategie E Criteri Di Interpretazione*. Roma: Carocci.

Bonnafus, S. (1998). "El análisis del discurso político",. *Primer Simposio Internacional de Análisis del Discurso*,. Madrid.

Bucheli, V. (2006). *Herramienta informática de vigilancia tecnológica para análisis socio-cognitivos de comunidades científicas*. Bogotá: Universidad Nacional de Colombia.

Cabrera Varela, J. (1987). Una aplicación informática para el estudio de casos. *Revista española de investigaciones sociológicas*, ISSN 0210-5233 , Nº 40 , 161-178.

Eagleton, T. (1997). *Ideología*. Barcelona : Paidós.

Enríquez Ruiz, G. (12 de Diciembre de 2009). *Proclama tu tierra. net Norte del cauca*. Recuperado el 6 de Octubre de 2010, de <http://proclamantortercauca.blogspot.com/2009/12/celebracion-de-la-independencia.html>

Etxeberría, J., García, E., & Rodríguez, G. (1995). *Análisis de datos y textos*. . Madrid: Ra-Ma.

Feldman, A. (2002). La integración del Análisis Textual a los estudios de Calidad de Servicio y Satisfacción del Cliente. En A. Morin, & P. E. Sébillot, *6èmes Journées internationales d'Analyse statistique des Données Textuelles*. (pág. volumen 1). Saint Malo.

Giménez. (1983). *El análisis del discurso político-jurídico*. México: UNAM.

Höppner, F., Klawonn, F., Rudolf, K., & Runkler, T. (1999). *Fuzzy cluster analysis. Methods for classification, data Analysis and image recognition*. John Wiley & Sons.

Ide, N., & Véronis, J. (1998). "Word sense disambiguation: The state of the art" . *Computational Linguistics*, 24(1), , 1-41.

Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.

Kodratoff. (1999). Knowledge Discovery in Texts: A Definition and Applications. *Proc. of the 11th International Symposium on Foundations of Intelligent Systems (ISMIS-99)*. , (págs. 31 - 58).

Larocca, J., Santos, A., Kaestner, C., & Freitas, A. (2000). Generating text summaries through the relative importance of topics. *Proceedings of 7th Iberoamerican Conference on Artificial Intelligence*, , (págs. 300-309).

Lebart, L. (1986). Analyse statistique des réponses libres dans les enquêtes par sondage. *Revue Française du Marketing*, 109.

Lebart, L., & Salem, A. (1998). *Analyse Statistique des Données Textuelles. Questions ouvertes et lexicométrie*. . Paris: Bordas Dunod.

Lebart, L., & Salem, A. (1994). *Statistique Textuelle*. Paris.: Dunod.

Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle* . Paris: Dunod.

Lebart, L., Salem, A., & Bécue, M. (2000). *Análisis Estadístico de Textos* . Madrid: Milenio.

Lefort, C. (1986). *The political forms of modern democracy: Bureaucracy, democracy, totalitarianism*. Cambridge: Polity Press.

Maingueneau, D. (1989). *Introducción a los métodos de análisis del discurso*. . Buenos Aires : Hachette.

Mendieta, E. (2007). "Hacer vivir y dejar Morir": Foucault y la genealogía del racismo. *Tabula Rasa*, 137 - 152.

Montes-y-Gómez, Gelbuhh, & López-López. (2001). A Statistical Approach to the Discovery of Ephemeral Associations. *12th International Conference on Database and Expert Systems Applications* (págs. 491-500). Munich: Springer-Verlag.

Nürnbergger, A., Klose, A., & Kruse, R. (2001). Clustering of document collection to support interactive text exploration. *Proceedings of 25th Annuals Conference of the Gesellschaft für Klassifikation*. , (págs. 291-299).

Pecheux, M. (1978). *Hacia el análisis automático del discurso*. Madrid: Gredos.

Pérez Hernández M, C. (2002). Explotación de los corpóra textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística Española (ELiEs)*, ISSN: 1139-8736, , Volumen 18.

Polanco, X. (2001). Experiencia del INIST en la producción de indicadores infométricos. Indicadores y Vigilancia científica y tecnológica,. *Ponencia Taller Andino de Indicadores de Ciencia, Tecnología e Innovación*. Bogotá: CAN, República de Francia, COLCIENCIAS, OcyT, RICYT, PUJ.

Polanco, X. (2002). Transformación de la información en conocimiento y del conocimiento en decisiones estratégicas. *Unité de Recherche et Innovation, Institut de l'Information Scientifique et Technique (INIST)* .

Reinert. (1993). "Les "mondes lexicaux" et leur logique ". *Langage et societe* , 5-39.

Robin, R. (1973). *Histoire et linguistique*. París: Armand Colin.

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, And Retrieval Of Information By Computer*. Reading, Massachussets : Addison-Wesley.

Sánchez, A., & Almela, M. (2006). Formalización de las correspondencias entre acepciones y contextos sintagmáticos en español e inglés. *Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística*. León: Universidad de León, Dpto. de Filología Hispánica y Clásica.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press. .

Sperber, D. &. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. . Oxford: Blackwell Publishers.

Thompson, J. (1993). *Ideología y Cultura Moderna*. México: UAM Xochimilco.

Verón, E. (1987). La palabra adversativa. En V. y. otros, *El Discurso Político*. Buenos Aires: Hachette.

Villaseñor, L., Aurelio, L., & Montes, M. y. (2003). Tratamiento Automático De Textos En Español. *Estudios de Lingüística Aplicada, diciembre, año/vol. 22, número 038* , 145-166.

Weber, M. (1978). *Economy & society: An outline of interpretative sociology*. Berkeley: University of California.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA : Addison-Wesley. .

ANEXOS

ANEXO: CORPUS TRABAJADO

¿Historia?

La historia es redención, no expiación

El más sabio padece errores y el más justo practica el pecado, me decía un amigo refiriéndose a los Estudios sobre la vida de Bolívar, porque las pasiones ofuscan el entendimiento y socavan la voluntad". Y así sucede: esos Estudios, cuyo autor es dueño de una penetración intelectual y de un acervo de conocimientos no comunes, aparecen saturados, desde el principio hasta el fin, de una atmósfera de odios y aversiones injustificables. Leyendo esa obra, le parece a uno estar en el cubil de una fiera, que así devora con impaciencia cadáveres de hombres como despojos de perros. Y es que la inteligencia privilegiada y la rectitud ética fueron entradas a saco, primero por la credulidad, más tarde por el prejuicio y finalmente por la pasión inconfesable y funesta. En la misma introducción, digna por su vuelo de una causa noble, que no de la absurda y antipatriótica que movió la pluma del doctor Sañudo, hay conceptos que nunca cuadrarían a un predicador de moral y a un maestro de filosofía. Hace años oímos que el doctor Sañudo pedía a Dios el don de la vida hasta publicar un libro sobre Bolívar, que sería una revelación. Dios, en su designio inescrutable, le oyó el ruego. El libro ha aparecido, para escándalo e indignación del patriotismo; y el doctor Sañudo ha muerto. Ha muerto irremediamente; de muerte natural y definitiva, sin derecho siquiera a esperar la resurrección de la carne. Porque cuando los hombres importantes mueren para la opinión pública, nunca vuelven a triunfar ni a vivir en ella. Se curan las crisis del prestigio que provienen de causas pasajeras; pero la muerte es muerte, y sobre el cuerpo de sus víctimas cae una losa enorme que jamás se levanta.

Y decimos que el doctor Sañudo ha muerto de muerte natural, porque el enemigo de la independencia, debía ser lógicamente el enemigo de Bolívar; y porque el insultador del Padre de la Patria, debe la Patria ejemplarmente castigar, alejándolo de todas sus lógicas manifestaciones.

Más de una centuria llevamos de vida autónoma. La paz y la guerra, la guerra y la paz han envuelto a la Nación. El bien y el mal se han apoderado de ella. La garra extranjera se ha cebado en sus entrañas. Y en medio de esa muy humana lucha de fuerzas encontradas, sólo la memoria de los fundadores de la República ha permanecido nítida y estimulante. Le tocaba al doctor Sañudo la triste pretensión de mancillarla. Mas, para honra de la tierra y glorificación del inmortal caudillo, no lo ha conseguido. Porque son tan grandes los desatinos que afirma y tan de bulto las necedades que profesa, que todo lector desapasionado logra pasar sin mareo por sobre las olas aparentes de un capricho y de una injusticia sin nombre.

Después de intentar rápida e inútilmente la refutación de algunos sistemas de filosofía de la historia, en verdad refutables pero dentro de los cuales no arma -según su léxico- la figura raquítica y malvada que, contra toda moral y toda sana crítica se ha ideado de Bolívar, funda la historia en una base de rara expiación, en un sistema original trabajado especialmente para hacer deducciones contra el Libertador, de hechos que arguye comprobados o posibles, con maliciosas e infantiles afirmaciones.

Con una modestia genial, propia de los hombres que se fabrican una soberbia torre de marfil y la suponen inexpugnable, dice ser la opinión fundamental de la historia que él sostiene, la generalización de la sublime e incomparable de Bossuet.

En casi tres siglos no había aleteado el espíritu fuerte que viniera a complementar lo que el Águila de Meaux dejara falto y defectuoso. Le esperaba al doctor Sañudo la suerte de definir lo que, según el mismo, hasta hoy nadie ha definido y que fue en muchos años y sigue siendo desvelo y Preocupación de insignes filósofos y profundos historiadores.

Que la Historia, como Bossuet lo expresa, gira al rededor de la redención, es una hermosa y trascendente verdad. Pero el grande orador no confundió ni podía confundir la Redención con la expiación. El fin amoroso y Providencial de Dios al mandar a su Hijo Santísimo al mundo para que tomara nuestra débil carne mortal, era redimir al género humano de la esclavitud del pecado y del demonio. Y el medio, también providencial - e inescrutable, de esa redención, - era echar sobre los divinos hombros inocentes de Jesucristo, la carga de pecados de la humanidad, para que - optándolos El única víctima satisfactoriamente propicia a la Deidad- viniera la redención al hombre.

La Redención, pues, era el fin; la expiación, el medio. De manera que Ira la Historia de la humanidad al rededor de la Redención, es decir: del perdón del amor y de la infinita misericordia de Dios.

Así, según el pensamiento de Bossuet, la expiación de Jesucristo produce consecuencias divina y eternamente favorables a la humanidad; según el pensamiento del doctor Sañudo, la expiación es necesidad inevitable que va siempre cargada del pecado! y la catástrofe; según Bossuet, ira la Historia al rededor del amor y de la vida; según el doctor Sañudo, i historia gira al rededor del odio y de la muerte.

Y donde se nota la parcialidad del historiador respecto de Bolívar, es en sostener su curioso principio de expiación, sólo desde el Libertador para delante y no aplicarlo desde el Libertador también hacia atrás. Porque conocía que así la madre España quedaría envuelta en una maligna y espesa sombra de pecados y de crímenes que la historia—por medio de Bolívar—debería expiar. Esto, que no afirmamos nosotros, es lógico entro del sañudo sistema de la expiación.

Este principio, que puede ser una pequeña parte de la explicación de la listona en lo que a las naciones se refiere, desfigurado, exagerado por e1 doctor Sañudo, viene a desenvolverse en un fatalismo invencible, en un pesimismo desconcertante y en simplismo histórico que desacredita lentes capaces de conocer las complejidades fenoménicas del espíritu y de la moral, porque cabe en ojo estrecho y miope, en que no pueden reflejarse el universo en su enorme complicación armónica y lo que más allá del universo y -dominándolo- existe.

Metidos un instante en el expiatorio sistema del doctor Sañudo, se nos imagina la Historia un sombrío campo de tragedia sobre el cual se ciernen las alas gigantescas de Esquilo, de Shakespeare y de Ibsen. Y casi anonadados, inquirimos: ¿no será el autor de los Estudios sobre la vida le Bolívar un formidable dato histórico de la ley de la expiación?

Pensando además en las varias y complejas causas que pueden –sobre todo por el aspecto humano- producir la Historia, creemos que el doctor Sañudo, asignándole solo una y todavía a su modo, la expiación, yerra notablemente.

Si en lo físico, con sus leyes inevitables, son muchas las causas de un fenómeno; en lo moral, especialmente en lo colectivo -abstracción hecha de la indiscutible y necesaria intervención de la Providencia Divina- no lo son menos. A este propósito o por esta aseveración, recordamos la frase de Le Bon: "une des caractéristiques de mentalité primitive est d'attribuer des causes simples aux phénomènes"

Otro gravísimo defecto de crítica del doctor Sañudo, es analizar al Libertador fuera de la Historia; es decir, aislado del medio y de los acontecimientos que lo rodearon; ver en sus acciones sólo un proceso radicalmente individual de conducta y no el resultado muchas veces, casi siempre, de las circunstancias, que encaminaban su libre albedrío -sin destruirlo- por senderos determinados; sacarlo del campo donde la Providencia lo había colocado como un libre instrumento de su voluntad eterna y suponerlo una celebridad sin fines de trascendencia.

Por otra parte, olvidó el doctor Sañudo que —respecto de los grandes hombres que influyen en la marcha de la humanidad— no se tienen en cuenta los hechos o rasgos puramente individuales, pequeños, insignificantes, buenos o malos, por la sencilla razón de que ellos no forman la Historia. La vida íntima de esos hombres es alimento de la curiosidad biográfica o de la novela recreadora; no materia de tranquilos y serios historiadores. Un célebre crítico de Cousin dice: "de la théorie des peuples, Cousin passe á celle des grands hommes dans les queis ils se personnifient. Un grand homme, en effet, représente, sous une forme particulière et individuelle, l'esprit general de sa Nation. Ce n'est donc pas sans raison que les historiens mettent ordinairement les grands hommes en pleine lumière, et laissent dans l'ombre les masses confuses dont ils ont été les personifications plus ou moins éclatantes. Il y a des grands hommes toutes les fois qu'il y a dans la société de grandes idées qui cherchent leur expression. Ils naissent, croissent, déclinent et meurent avec elles, tant il est vrai que c'est á elles qu'ils empruntent leur vie et leur puissance. Un grand homme peut sans doute avoir de tres petits cotes; mais ce n'est pas par ceux-la que l'humanité le considere: c'est par ceux qui constituent ses rapports avec elle et qui font sa grandeur. Les traits individuels et vulgaires, qui le caractérisent, peuvent être relevés par les mémoires et les biographies; les traits généraux et nobles sont seuls burinés par l'histoire. Celle-ci néglige ses mobiles, souvent égoïstes et misérables; elle s'attache á ses actions, pour peu qu'elles aient contribué soit á la prospérité d'un peuple, soit á celle de l'espèce". Y en otra parte: "Il a raison de dire que les grands hommes representent l'esprit d'une Nation á un moment donné de l'histoire".

Y sobre la base de que Bolívar fue un hombre sujeto como todos a las pasiones, sin constituir, ni muchísimo menos, el monstruo nacido de la mente del doctor Sañudo, hay que advertir que las pasiones individuales nunca tienen fuerza de transformar la Historia y no influyen sino en sus accidentes, pero no en su esencia. El mismo crítico francés citado, refiriéndose al sistema de filosofía de la Historia del gran Jouffroy, dice: "Il est á remarquer cependant que l'influence des passions individuelles, si haut place que soit celui qui leur obéit ne peut modifier que les événements secondaires: les événements importants lui échappent". Y el mismo Jouffroy,

bellamente: "En définitive, la passion n1 agit qu' á la surface de l'histoire de peuples: le fond appartient aux idées; il en est effet logique et l'expression puré".

Y eso es natural, porque la humanidad -como tal- y las diferentes sociedades, se guían para su felicidad, que necesariamente buscan, no por los sentimientos aislados de cada hombre, sino por su instinto colectivo del bien.

De suerte que no es Bolívar la causa de las penas, de los sacrificios, ni de los errores de la República -como infantilmente lo asevera el doctor Sañudo, aplicando a su principio de la expiación el más vulgar de todos los sofismas: post hoc, "ergo propter hoc".

Estas pequeñas líneas, sobre la introducción del libro especialmente, por lo pronto.

Que si Dios no dispone otra cosa, seguiremos analizándolo más despacio, tanto en su fondo filosófico y moral como en su valor histórico y en su tendencia e intención políticas, que en esta ocasión le ha sido a su autor imposible disimular.

Un indio patriota

Aún vive la historia

Todos tenemos derecho de hablar cuando son conculcados los hombres que se quemaron por la Patria en que vivimos. Los esbirros españoles que mataron en su origen a los descendientes de Manco Capac, Atabaliba y Guascar y Montezuma, no son dignos de preseas y triunfos honoríficos, cuando se recuerda la espada brutal que acuchilló a nuestros aborígenes de toda la altiplanicie andina. La conquista, mirada a la luz de la razón, no fue sino el aplastamiento de la fuerza bruta contra el derecho de la existencia de los pueblos de las Indias Occidentales. Cuando el grito de la Independencia surgió de los pechos de los criollos contra el dominio de los españoles, no fue sino la protesta contra todo lo inicuo y bajeza de nuestros amos los señores españoles. Los cadalzos levantados contra Caldas, Camilo Torres, Pombo y Jorge Tadeo Lozano, claman venganza contra Boves, Morillo, Morales y Monte Verde y contra el sanguinario Sámano, el diablo del terrorismo. Que la España sepa una vez por todas, que los criollos americanos, llegados a la mayor edad, tenían pleno derecho al Gobierno de sus pueblos; y que tratar de ahogar la Independencia por medio de la fuerza, fue el acto oías bárbaro del despotismo español. Si España hubiera obrado como Reina y Señora de sus pupilas las colonias suramericanas, entonces no habría para qué entrar a analizar la fermentación odiosa de la Independencia americana. Que la Independencia necesitaba el hombre providencial para su coronamiento, no hay para qué discutirlo. Que Nariño hubiera sido el hombre, no hay para que ponerlo en tela de juicio. Que lo fue Bolívar el héroe más grande de la América, la Historia no lo discute. Que tuvo errores muy grandes e hijo de Adán como los demás hombres, tampoco hay que negarlo. Que descendía de Condes y Marqueses y vizcaíno, Bolívarjáujauregui, —pradera del molino— y noble en toda la acepción de la palabra, ya lo dice Unamuno, sabio español: "Mejor que nadie acaso conocía a Bolívar su más noble contendor —que no fue el virrey Sámano, ni el virrey Montalvo, ni el virrey La Serna, ni el francés Canterac, sino el general español don Pablo Morillo, —y decía de él: "Tiene de su noble estirpe española rasgos y cualidades que le hacen muy superior a cuanto le rodea". "Y, sin duda, muy superior a los que llevaran sangre de Manco-Capac, a cuyos hijos remachó las cadenas Pizarro, aquel Pizarro mucho mas hermano de Bolívar que el inca". Que en Bolívar brilló la justicia el mismo Unamuno, español, lo va a decir: "En mi vida de don Quijote y Sancho, he escrito: don Quijote castigaba, es cierto; pero castigaba como castigan Dios y la naturaleza, inmediatamente, lo cual es naturalísima consecuencia del pecado. Así Bolívar. Fusiló a propios y a extraños pero jamás con ensañamiento. Su justicia, como la de don Quijote, era rápida y ejecutiva. Boves lo derrota en La Puerta, y hace una carnicería de las suyas: Bolívar fusila inmediatamente a ochocientos prisioneros. Piar, su teniente se insubordina, huye del ejército y trastorna el orden en momentos angustiosos: Bolívar lo hace aprehender, juzgar y fusilar. Lo propio hizo con Berindoaga, ministro de Guerra y traidor en el Perú. Lo mismo con Vanoni el único de los realistas vencidos, a quien fusiló en el campo de Boyacá, en 1819 porque siendo oficial suyo lo había traicionado en 1812, haciéndole perder el castillo de Puerto Cabello" "La Justicia sola es la que conserva la República", decía (Cartas del Libertador: O'Leary, volumen XXX).

Que Bolívar fue cruel: lo contrario dice Blanco Fombona de la guerra a muerte en 1813: "No es la crueldad fría de los corazones felinamente tiernos, débiles, dice; es el rugido de desesperación y dolor de los corazones generosos, pero recios. Los reveses hacían temible a Bolívar y el éxito lo

hacia magnánimo." Que el Libertador no tenía las charreteras de un verdadero militar. Esta cuestión tan manoseada no admite réplica. Mientras los españoles maniobraban, vacilaban, calculaban, guardaban las acostumbradas avenidas de aproche, Bolívar los sorprendía concentrando una fuerza superior sobre el punto donde menos se esperaba el ataque, introduciendo la confusión y batiendo las tropas en detalle. Por eso las acciones de Bolívar en el campo de batalla no pueden ser expuestas de manera que impresionen tanto como las de otros generales menos notables." El gran historiador español Opisso, dice hablando de la batalla de Boyacá; "Esta batalla de Boyacá, era después de la de Maipó la más trascendental ocurrida en la guerra de Sud América, pues daba a los patriotas la preponderancia en el norte de este continente, como Maipó se la había dado en el Sur. Nueva Granada fue toda de los patriotas, y Morillo quedó aislado en Venezuela." Que Bolívar fue el autor de la guerra a muerte? Allá va la historia que lo desmiente. Dice Opisso, español, lo siguiente: "Achacóse por algunos al marino don Domingo de Monteverde, tan valeroso como intransigente, la culpa de haber inaugurado el sistema terrorista, a lo cual se añadió la aparición de partidas sueltas realistas que procedían a usanza de nuestros guerrilleros de la guerra de la Independencia y estaban por lo general compuestas de gente non sancta, pero cuyos servicios no se podían desdeñar, con lo cual adquirió la lucha el carácter terrible que revistió por tantos años." "Guerra espantosa, escribe el historiador D'Espagnat, guerra a cuchillo, guerra para acabar más épica y desesperada por parte de los rebeldes, más rencorosa y furibunda en las filas de los españoles, guerra inexpiable también con sus esplendores y sus salvajadas, sus páginas a lo Floro a través de la concisión realista de los relatos de Polibio, sus rasgos dignos de los olímpicos héroes de Troya, y otros que parecen resucitar los horrores inventivos de las batallas de Almirante".

Que los españoles fueron crueles hasta la barbarie de pueblos salvajes: citaré otra vez al historiador español. "Rivalizaban, pues, en sanguinaria furia unos y otros. Figuraban entre los más terribles jefes españoles Cerveriz, Antoñanzas, Zuázola, Boves, Fray Eusebio de Coronil, Capellán de Monteverde, Rósete, Yañez, todos ellos muy valientes, pero cerrados a la piedad enloquecidos por el odio. El uno se jactaba de no perdonar a ningún patriota prisionero; el otro hacía matar a niños y viejos, y despanzurraba a las mujeres en cinta; éste se entretenía en cortar las orejas a los cadáveres para que su gente las llevase clavadas en el sombrero a guisa de escárpela; aquel se gozaba en torturar a los que caían en su poder. El nombre de "Estandarte de la Muerte," como se llamaba a la bandera de Boves da idea de lo que sería la división infernal que mandaba". Por consiguiente, no alabemos a los españoles, para denigrar a los criollos; esto vale escupir al cielo, para que nos caiga en la cara.

Seguimos con que Bolívar aprisionó a Miranda. La traición fue la estrella de Miranda, dicen los mismos historiadores españoles. La capitulación que hizo Miranda, con Monteverde, teniendo aquél 5000 hombres listos para proseguir la guerra, indignó al pueblo y lo acusó de traición. Amén, de que cuando los patriotas que emigraron a La Guayra, tuvieron conocimiento de que Miranda se alojaba en las habitaciones del Comandante de la plaza, que era entonces don Manuel María de las Casas, ya a la sazón en connivencia con los españoles. Entonces, sabiendo esto Bolívar, y con el conocimiento que tenía de la traición de Miranda, lo hizo aprehender y conducir al fuerte de San Carlos. No hay que hacerse ídolo de los septembristas. No se los puede alabar en ningún caso. La Iglesia condena y reprueba esos actos como contrarios a todo derecho y a toda moralidad. Nunca por autoridad privada y particular es lícito matar al Príncipe, aun cuando sea tirano, mucho más cuando sea Dictador, que es muy distinto. Alabar a los septembristas en concreto, tal como

obraron esos cobardes, buscando las sombras de la noche y con puñal en mano para intimar al Libertador, y asesinando a la guardia nacional, es prohijar uno de los hechos más escandalosos de barbarie y de asesinatos perpetrados bajo la capa de derechos conculcados. Que Bolívar fue el mayor revolucionario del país? No es cierto: estaba en paz y pacífico el Gran Libertador. Por consiguiente, era intocable e inmune su vida; cara existencia, que en mil combates respetaron las balas españolas, y solamente la mano traidora y villana de los descastados de Colombia, quiso manchar a la Nación con el negro baldón de parricidas!

Las guerras consideradas en particular y en el desenvolvimiento sucesivo de las batallas en acción, como también en los tiempos en que se dieron, es un asunto que hay que considerarlo imparcialmente y sin ninguna clase de venganzas y odios. Los Generales que tienen el mando supremo en tiempo de campaña, deberían ser más humanos y obrar en todo con rectitud de conciencia; pero, desgraciadamente, resulta todo lo contrario; vencedores y vencidos se juran odio a muerte, y la venganza de sangre se hace sentir como una ola de fuego contra los prisioneros, vencidos y rendidos. El Derecho de guerra prohíbe poner celadas y emboscadas cuando no hay campaña, como resultó en el Sur de Colombia con don Julio Arboleda. Aquí, en esta parte, no había combates: por consiguiente, aun cuando don Julio venía escoltado, no se sigue que era lícito asaltarlo y asesinarlo como lo hizo López. Esa clase de bandidos no son beligerantes, sino asesinos en todos los idiomas del mundo. Lástima grande que el doctor Sañudo, hombre de letras y jurista renombrado; gran historiador y filósofo de gran talla; matemático insigne y políglota versado en muchos idiomas; gran pedagogo y modestísimo en su vida pública y privada, lástima, decimos, el que en sus Estudios sobre la vida de Bolívar, no haya brillado la imparcialidad al hacer el recuento de las hazañas y proezas del Libertador.

Alexo Corfú

¿Historia?

La Historia es previsible

Negaba el doctor Sañudo la posibilidad humana de prever la historia. Pero, inconsecuente y urgido a veces por la lógica y la experiencia, que, a su pesar, se le imponen, admite ese poder previsor, partiendo únicamente de lo que él con gracia apellida descarríos de los Estados y Naciones.

Reafirma aquí su estrechísima y asfixiante doctrina de que la historia humana no es más que un zurcido de culpas y un zurcido no men6s espantoso de castigos. No pueden los Estudios imaginarse a la humanidad sino pecando, vencida y agobiada por una inclinación fatal; y a sus Estados y Naciones, delinquiendo eternamente, víctimas convulsas y epilépticas, sin poder ofrecer al mundo otra cosa que sus faltas y expiaciones.

A esta teoría, que ahora podría denominarse del descarrío, le busca el enemigo de Bolívar objeciones infantiles que después aplasta con indiscutibles postulados de filosofía social. Nos parece en tales momentos un jugador de ajedrez que se divierte consigo mismo, dirigiendo el ataque de ambos escuadrones de fichas y cansando hasta el extremo del bostezo a los inocentes y desocupados espectadores.

¿Quién no ha estudiado o leído siquiera una vez en su vida la famosa, atomística, mecánica y disolvente exposición de Rousseau respecto de la sociedad? ¿Quién no sabe que esa endiablada tesis así conduce a la anarquía como arrastra a los más feroces despotismos? Recordamos a un viejo profesor, que para prestigiar principios y conclusiones, nos decía: Santo Tomás, Kant y los mejores filósofos del mundo, siguen mi teoría. Los Estudios —no para que lo sigan— sino para darse el lujo de un contendor formidable, declaran opositor de su sistema expiatorio, a nadie menos que a Rousseau. Cuando aun dentro de la concepción del ginebrino misántropo, el Estado, mejor dicho el Gobierno, debe ser la expresión no de todos los individuos numéricamente considerados, sino de la mayoría. Confunde el injusto panfleto en las ideas de Rousseau el origen de la sociedad civil y su gobierno. En el origen vienen todas las voluntades particulares a la Constitución del pacto social; no así en el gobierno, que —por lo que se nota— es para el doctor Sañudo el pecador, causa de la expiación de los pueblos. No cabe, por esto, ni aun en el falso, anarquista, despótico y disolvente sistema de Rousseau, la objeción de que no sería posible la filosofía de la historia de un pueblo o de un Estado, porque debiera entonces buscarse de todos y cada uno de sus miembros.

La dificultad aquella parece el juego de palabras de los mamotretos insoportables que indicaron en los siglos catorce y quince la decadencia de la escolástica; escolástica que ya no era aquella robusta y grandiosa de Alejandro de Hales, Alberto Magno, Tomás de Aquino, Rogerio Bacon, Lulio, S. Buenaventura, Durando y especialmente Escolto, el entendimiento quizá más fuerte y bello con que Dios haya ornado la tierra.

Es verdad que existen faltas de la sociedad distintas de las faltas de los individuos considerados particularmente. Más todavía: según lo expone con envidiable claridad el P. Weiss, esas faltas sociales, son distintas de la suma de las faltas de todos los individuos que la forman. Pero es un gravísimo error sostener que siempre los pecados de los jefes, lo son de los Estados y Naciones. Lo

comprendió así el doctor Sañudo y arregló su concepto pasándose de las responsabilidades a la materialidad de las consecuencias, diferenciando pecados de los individuos en colectividad y pecados de sus jefes o rectores.

Luego, sin el cuidado y prudencia que competen a un hombre que se lanza a exponer cosas de tanto peso e importancia, expresa el doctor Sañudo, hablando de la sociedad civil, que "delitos que hombres aisladamente no los cometerían y aun se horrorizarían de pensarlos, unidos los cometen sin dificultad" y que "hombres eminentes se anulan en un Congreso, y votan leyes disparatadas que aislados jamás aprobarían". Sin duda, quisieron los Estudios demostrar con eso la diferencia no solo cuantitativa sino cualitativa entre la sociedad y sus miembros; pero esto debe expresarse claramente y no de manera confusa que puede convencer de serios errores a un lector impreparado. El flaco consiste en fugarse —como si tal— de la reunión que es Sociedad o Estado a otras que—aunque pudieran constituir entidades morales y hasta personas jurídicas— por sí solas no se confunden con aquéllos.

Y donde echan de su lomo escama los Estudios, es en la aplicación política que —traída de los cabellos— hacen de las transparentes y rectilíneas afirmaciones de Aristóteles y Santo Tomás sobre la constitución de las sociedades civiles. Es curioso el argumento: la sociedad difiere de la suma de los individuos que la forman no sólo cuantitativa sino cualitativamente; luego los partidos permanentes en los Estados son peligrosos, "porque sus secretarios, por instinto, aprueban el mal y desestiman el bien, sin que tengan conciencia de su defección moral". Nos recuerda este ingenioso modo de raciocinar, aquella frase, no sabemos si de una novela vernácula, que dice: "era de noche y sin embargo llovía."

Nosotros pensamos, por el contrario, que los partidos políticos permanentes son una consecuencia muy natural de la sociabilidad civil humana, que dentro de la misma variedad de opiniones sobre política y administración, toman posible el gobierno por medio de las mayorías. Dada la imposibilidad moral de una opinión o de un consejo unánimes ¿qué resta? Sólo el querer o la opinión o el consejo de aquellas mayorías ¿Y éstas, cómo pueden existir y demostrarse prácticamente? Sólo por medio de los partidos políticos, que no son otra cosa que principios en acción que viven luchando por imponerse y constituirse en gobierno y administración pública.

La teoría de los Estudios contra los partidos políticos lleva necesariamente a la anarquía, porque así se requiere para las decisiones gubernamentales la voluntad de todos y cada uno de los individuos, lo cual es imposible. Se ve que el doctor Sañudo, no maliciosamente, aplica a estas horas al gobierno de la sociedad civil, la idea atomística del pacto social que—por otra parte— rechaza racionalmente con vehemencia. Contradicción y más contradicción.

Profesa el autor de los Estudios respecto de los partidos una especie de eclecticismo indiferentista en política, que parece haberlo practicado durante su vida, aunque sí se le alcanzan a notar ciertas inclinaciones y simpatías. Para nosotros—humildes luchadores de ideas—el eclecticismo filosófico es el temor del entendimiento; y el eclecticismo político, la cobardía de la voluntad.

Para un espíritu resbaladizo, sin firmeza filosófica ni orientación política definida, ha de presentar pocos rasgos de simpatía, quizá ninguno, la magna figura afirmadora y rotunda de Bolívar, afirmadora en el fondo de las ideas, y más afirmadora, si cabe, en la realización del pensamiento. Y

no se objete que el mismo Libertador clamaba —en los umbrales de la muerte— porque cesaran los partidos y se consolidara la unión. No demostraba con esos anhelos fervientes el gran guerrero, la necesidad de acabar con los partidos políticos, que apenas comenzaban a iniciarse o vislumbrarse. Comprendía que hay un centro a donde pueden y deben confluír todas las fuerzas nacionales por opuestas que parezcan: el bien y la salvación de la Patria. Y para ese bien y esa salvación ansiaba Bolívar que desaparecieran las exageraciones partidistas personales y se efectuara una unión, dentro de la cual podrían moverse libremente y en paz todas las nobles actividades de la República.

No es exacto que los miembros de un partido político, por instinto, aprueben el mal y desestimen el bien. Sucede al contrario: el instinto moral— si así puede hablarse—que no desaparece en los movimientos e instituciones colectivos, llama e inclina al hombre hacia el bien, porque así es su naturaleza. Y sólo cuando la confusión de ideas y la efervescencia de sentimientos son muy grandes, llegan a no respetar los miembros de una sociedad los dictados de la conciencia, que siempre los impulsa a la verdad y al bien. Entonces obran contra lo que el doctor Sañudo denomina instinto.

Empeñados los Estudios en lucir su sistema expiatorio y llevarlo a todos los extremos para deducir después la responsabilidad de Bolívar en todas las desgracias de Colombia, considerándolo— contra toda verdad y justicia— un terrible bellaco y delincuente afirman ser justo que recaigan en los súbditos las consecuencias expiatorias de los pecados de sus jefes. No acude — expresa el doctor Sañudo— para llegar a ésa conclusión, a Lombroso, Von Listz ni otros positivistas o criminalistas sociales. Hace como un sabio: ¿que tienen que ver Lombroso y sus secuaces con la ley de la expiación? Nada, absolutamente nada. Al revés: para esos penalistas, el delincuente debe carecer de responsabilidad; y por lo tanto, ni para él ni para nadie puede hablarse de expiación o pena por culpas o pecados. El derecho penal positivista se parece a la ley de expiación del doctor Sañudo en lo que la geografía del Indostán se parece a los Estudios sobre la vida de Bolívar.

Va el doctor Sañudo a apoyar su injusta conclusión en otro principio que es necesario explicar y que los Estudios establecen de manera ilimitada y absoluta: la solidaridad de los actos humanos como opuesta al principio de la lucha por la vida, de Darwin. ¿Qué es la solidaridad así en el terreno privado como en el colectivo o público, desde el punto de vista de la ética? Creemos no estar equivocados respondiendo que consiste en la responsabilidad contraída de manera mancomunada. Mancomunidad que puede ser expresa o tácita. Tácita casi siempre que se trata de movimientos colectivos sociales. Todo el mundo sabe que la responsabilidad tiene por primera base racional la imputación, es decir, el hecho de ser una persona causa o autora del acto que se le imputa.

Y como no siempre los pueblos o los súbditos son socialmente coautores, cómplices, toleradores o instigantes de las faltas de sus jefes, que todo eso es base de mayor o menor imputación moral, síguese que no siempre se es pueden imputar, siendo entonces para ellos ningunas las responsabilidades. Y donde no hay responsabilidad, tampoco existe solidaridad. Esta implica un lazo moral entre los solidarizados; lazo que proviene de interior conocimiento de las cosas y del libre consenso de la voluntad. ¡Por qué ha de ser solidario de su gobierno radical, enemigo de la Iglesia, si pueblo católico de Colombia del 63 al 85, que no cejó en la lucha hasta llegar a su República una dirección conforme con las doctrinas religiosas que secularmente ha profesado?.

Indudablemente, hay circunstancias en que los pueblos, de modo positivo y directo o por miedo y abulia, se hacen socialmente cómplices de los pecados de sus jefes y gobiernos. En esos casos, sí hay una verdadera solidaridad y es justo que sobre ellos caigan también los castigos de las faltas que ayudaron a cometer, o que —sin protesta— dejaron perpetrar. Pero eso de pensar en que los pueblos, naciones y Estados, sean siempre solidariamente responsables de las caídas o pecados de sus jefes, rectores o gobiernos, es algo desatinado, injusto, monstruoso y opuesto a las más triviales reglas de la moral y del derecho. Y ya que citamos un ejemplo, en que no es el pueblo solidario con sus jefes, citamos —para ilustrar—otro en que sí hubo esa solidaridad: los Estados Unidos, arrebatando de Colombia a Panamá en 1903. Allí quiso el pueblo norteamericano, lo que Roosvelt, su director, quería; y le ayudó a llevar a la práctica el arbitrario y abominable anhelo. Distinta cosa es que los actos de los jefes y guaidores de pueblos, tengan necesariamente sus consecuencias y que éstas —no de modo formal sino materialmente— recaigan sobre los súbditos o simples ciudadanos. Pero eso se explica—no por solidaridad o responsabilidad conjunta de pueblos y jefes o gobiernos que no siempre existe; sino por la fuerza misma o alcance objetivo de esas consecuencias, independientemente de la idea de responsabilidad de los primeros. ¿Puede alguien sostener que la madre y los hijos que sufren de un déspota la muerte del esposo y del padre inocente; puede alguien sostener, sin declararse un monstruo de iniquidad, que esa madre y esos hijos están solidarizados con aquel déspota? Que responda el sentido común.

“Ejercemos rectamente nuestro derecho y sin pensarlo beneficiaremos a otros”. Estas dos palabras de los Estudios, sin pensarlo, comprueban nuestras afirmaciones. La solidaridad requiere pensamiento o ciencia personal o colectiva de una cosa o de un fin y decidida voluntad.. “El bien o el mal que practicamos redunde de hecho en bien o en mal de la sociedad”. Es otra proposición con que contradice el doctor Sañudo su misma tesis. De hecho, es verdad; no de derecho. Pero el simple hecho de sufrir la consecuencia de un acto de otro, no funda responsabilidad ni solidaridad éticas de ninguna clase. Es cierto —como indican los Estudios, que “el delincuente, atacando un derecho, ataca la convivencia de los demás asociados”. Eso indica no haber solidaridad entre éstos y aquél; porque si apareciese, no habría ataque alguno a la convivencia de los demás, que así convivirían con el criminal en la idea y comisión del delito.

Desaparece más todavía la solidaridad extrema como regla, cuando se trata de actos de simples hombres, cada uno de los cuales, aunque objetivamente sufriera consecuencias, carga con su responsabilidad, si la tuviera. ¿Por qué ha de ser el doctor Sañudo solidario de los crímenes del Hombre Fiera; y nosotros, del asesinato horrendo de los Cayapas? Y si la Santa Biblia dice “que Dios castiga los pecados de los padres en los hijos”, no declara que la Divinidad castigue a los hijos y nietos por los pecados de sus padres y abuelos.

El medio justo que nosotros sostenemos, atacando la solidaridad extrema de los Estudios, no es el sistema disociante de Darwin, ni el exagerado individualismo liberal, -porque se reduce a establecer, según lo indica la sana filosofía, la responsabilidad como base del problema de la solidaridad.

Vuelve el doctor Sañudo al primer pecado para afianzar su tesis de lo solidario extremo en la humanidad. Para ello declara una completa analogía entre esa culpa y cualquier acto humano de otros hombres que Adán. Hay en el asunto una diferencia enorme. Oígame lo que bella y profundamente dice Naville, que nos parece estar de acuerdo con el pensamiento de la teología y

de la Iglesia Católica: “Si se hubiese de admitir que un ser como uno de nosotros hubiera pecado, y que ese pecado hubiera sido imputado a otros seres, otros en el sentido absoluto de la palabra; si se hubiese de admitir que soldados enviados de lejos para reforzar una guarnición, serían tratados como culpables de un acto de sedición que hubiera tenido lugar antes de su llegada; esta idea se chocaría tan directamente con el sentimiento de la justicia, que la conciencia humana no querría siquiera examinarla. Pero ésta no es la enseñanza cristiana. La enseñanza cristiana encierra una afirmación que puede traducirse en estas palabras: el acto que perturbó el orden de la creación no es el acto de un individuo, en el sentido que damos actualmente a esta palabra, sino de un individuo primitivo que no sólo participaba de la naturaleza humana como uno de nosotros, sino en el cual, porque era primitivo, esa naturaleza se hallaba concentrada por entero, de tal suerte que podría llamársele el hombre humanidad. Sus actos reunían dos caracteres, que desde entonces han sido distintos: eran a la vez individuales y humanos, en toda la extensión de este último término. En realidad, la humanidad entera estaba presente en el que cayó, y que era su jefe, su germen y su manantial.” Nada agregamos a conceptos tan claros y compendiosos. La ley de la herencia, que en verdad existe, no es ilimitada y exige su explicación. La herencia física se impone: los hijos son carne y hueso de la carne y hueso de sus padres. No reciben los hijos en la sangre de sus padres sino lo que estos tienen: si enfermedades, enfermedad; si deficiencias, defectos; si salud, salud. La herencia moral nunca padece ni padecer puede tanta fatalidad. A lo más, pueden formarse con el elemento corporal ciertas inclinaciones o predisposiciones para determinados actos, hábitos y costumbres; predisposiciones o inclinaciones que nunca por sí mismos pueden vencer la libertad humana. Los defectos morales, que atañen a la voluntad, no pueden ser transmitidos directamente de padres a hijos; porque constituyendo esa una facultad espiritual del alma, es imposible que por generación pasen a los generados sus condiciones o cualidades. Además, el alma es creación directa e inmediata de Dios ex nihilo, sin materia o elemento previo. De donde se deduce que, propiamente en ella, al unirse al cuerpo, no hay gérmenes ni reflejos del alma de los padres, alma que es la sede de los defectos morales o de las buenas condiciones anímicas.

No negamos, antes convenimos, en que la íntima unión del alma con el cuerpo, unión que forma un todo sustancial, puede producir—como dijimos, inclinación o predisposiciones o aptitudes especiales, por ser el segundo el órgano natural de la primera. Pero admitir eso, no es pensar que los padres transmitan a los hijos sus defectos morales y que éstos inevitablemente los hayan de padecer.

Tampoco es exacto —por ser demasiado absoluto— que el instinto social “cubre de infamia a la descendencia de un delincuente, o de honor y gloria a la de un benefactor”. En primer lugar, porque no es amoldado a la justicia hacer responsables a los hijos de los delitos y culpas de sus abuelos; y en segundo lugar, porque la experiencia apunta que descendientes de malos hombres han sido reverenciados en la humanidad por sus virtudes, talentos y acciones; y al revés: nietos de antepasados ilustres han caído en el aborrecimiento de la sociedad por sus proceder y conducta. ¿No venera el cristianismo en sus altares a muchos descendientes de sus perseguidores, que lo aceptaron y por él murieron? ¿No recuerda con horror la historia a los hijos de reyes santos, que hicieron de su voluntad un tirano, de su trono un déspota y de su corte una cueva de corrupción abominable?

Y una cosa original, pero lógica en el doctor Sañudo, es hablar de las leyes de la solidaridad y de la herencia, casi únicamente para el mal; y casi abandonarlas en lo que al bien atañe.

Las frases de Weiss, dominico suizo, podrían confirmar la solidaridad extrema de los Estudios, si fuesen solas, si no tuviesen relación alguna con explicaciones anteriores y posteriores de su formidable apología. En el fondo, Weiss basa la solidaridad en las responsabilidades sociales, sin llevarla a esa línea en que desaparece la individualidad humana, perdida en la generalidad. Citamos en corroboración los siguientes párrafos del ilustre apologista en la misma obra y muy cerca de los que citan los Estudios: “Por consiguiente, aunque todos los individuos tomados en conjunto forman la actividad pública social, es necesario distinguir bien la actividad, la moral, la opinión de la sociedad de lo que hacen los individuos cuando obran en nombre propio: de modo que la sociedad, como unidad, como organismo, tiene su tendencia intelectual, su moral y su actividad propias.”— “En ese caso es necesario, sin embargo, distinguir la actividad social, la actividad del todo, aunque sea producida por la actividad de los hombres, de la actividad personal de todos los individuos”—“El crimen de un individuo, recordemos tan solo a Nerón y a Robespierre, es, por decirlo así, la flor, el resumen, el coronamiento de los pecados que la totalidad ha amontonado” “Las faltas de los hombres son faltas de la humanidad. -Por verdadera que sea esta última afirmación, necesita, sin embargo, ser explicada; podría ser mal comprendida, y de hecho lo es. Unos exageran la falta de la totalidad hasta eximir al individuo de toda culpa personal: según ellos, este únicamente sufre por el hecho de la corrupción general; otros no le dan importancia, creyendo que la corrupción de la totalidad, la llamada culpa hereditaria o pecado hereditario, no es más que el resultado de los pecados personales de los individuos, es decir, la suma y resumen de ellos. Pero indudablemente son erróneas esas opiniones.” Todo lo anterior está expresando que el sabio dominico acepta la ley de la solidaridad en las sociedades y la humanidad, sobre el fundamento inmovible de la idea de una responsabilidad propia de aquéllas, como tales, pero al fin responsabilidad.

Ese es el justo medio y por tanto la verdad. Dentro de ese ámbito se explica la frase de San Ambrosio, citada por el doctor Sañudo: “Regum lapsus, poena populorum.” Es decir: la caída, el error, el pecado, la falta de los reyes es la pena o sufrimiento de los pueblos. No dijo el Santo: regum lapsus, culpa populorum. Dijo pena, como consecuencia, no siempre como castigo o expiación propiamente dicha.

Naturalmente, el doctor Sañudo estableció como fundamento de su historia una solidaridad humana llevada a los últimos e inaceptables extremos, para remontarse con sobra de injusticia y falta de lógica, porque su sistema es ilógico e injusto, hasta el Libertador y hacerlo responsable de las desgracias posteriores de la Patria, atribuyéndole miras que no tuvo y modos de ser y obrar que nunca fueron.

Nosotros —en cambio— vemos en las páginas civilizadas y gloriosas de Colombia, el efecto natural de la soberanía que para ella conquistó Bolívar, solidario de la obra de libertad que ha purificado y engrandecido al mundo.

Un indio Patriota.

Los refutadores del Señor Sañudo

Por fin, después de leer tantos escritos, que deshonran el país, por ser sólo acervo de insultos sin prueba alguna, sobre los Estudios que publicó el señor Sañudo, gusta el ánimo encontrar en el número de octubre de la "Ilustración Nariñense," uno firmado por Régulo Santal, cuya cultura da a entender, y miope será, quien no lo vea, que corresponde a un oratoriano de muchas letras, y, lo que vale más, lleno de virtudes y de patriotismo. Es varón doctísimo y hermano de quien con esa ilustre Revista, da más honor a Pasto que el señor Sergio Elías Ortiz con sus cuentos, a no ser uno que, por su clásico estilo, parecemos no corresponderle; o el señor Sañudo con sus indoctos escritos hechos y por hacer.

Dio Régulo una regla para analizar esos Estudios, de veras correcta; pues si el señor Sañudo atribuyó a Bolívar, perjurios, falsificaciones y asesinatos y llamóle revolvedor de Colombia, quien se echa a cuentas el refutarle procediendo discretamente, debe probar, o que el archivo de O'Leary en que fundó esos cargos es apócrifo, o que el señor Sañudo lo alteró por completo. Porque miren ustedes, que eso de andarse a la pesca de reparos tan insignificantes como el de que ese señor copió el epígrafe de una obra de Luciano, traducida al francés, es oficio de ratería, ya q' no siendo inventor del griego, ni de esa obra, de algún libro debió sacarlo, y saber su traducción, ora porque allí constaba, ora porque la hizo mediante un diccionario, ora por que alguno se la enseñó; de modo que aún queda al crítico probar que el señor Sañudo no sabe nada de griego, para que algo valga su reparo. Y si dice que por qué no lo puso en francés o castellano, puede que el señor Sañudo le conteste que no le dio la gana, y porque sólo a un crítico que odia, se le puede ocurrir que es pedante quien cita un trozo en lengua extranjera. Acabamos de leer el Milagro del padre Juan Mir, en cuya obra puso varias frases en griego; más tuvo la fortuna que en la culta España, no le saliera un zoilo de adversario, a apellidarle escritor de pedantería.

Ni la verdadera crítica histórica sufre que sólo se apoyen los hechos en el decir de un historiador, y menos en Larrazábal, que se propuso defender a Bolívar en todos sus crímenes, aún en los más probados, porque quizás el señor Sañudo aduzca otro también; sino en documentos y eso comprobados como ciertos.

El señor Velásquez Herrera, sin convite, se metió en la danza, y tan bien supo decir, que de su escrito se deduce el famoso principio de que el que medita lejos del mundanal ruido, no llega a la verdad, sino quien se codea en ferias y bodegones; porque cierto, Santo Tomás, a quien sus discípulos llamaban, por su silencio, el buey mudo, y que vivió en el retiro de un claustro, es un escritorillo de por ahí; mientras que, ya lo creemos, si el señor Velásquez se da un paseíto entre los Motilones del Magdalena v. g., a los pocos días sale motilón, como cualquiera de ellos. ¡Tan bueno es su principio! Lo mismo puede acontecer a los escritores serranos de Ritos de Tumaco.

En la misma "Ilustración Nariñense" salieron dos artículos, firmados por un Indio, que algunos dijeron ser del doctor Belalcázar, lo que no nos parece porque el jamás tiene semejante confusión de ideas. Por uno, hasta se llama soberbio al señor Sañudo, por haber dicho que la ley de la expiación, aplicada a todo pueblo, es la generalización de la que dedujo Bossuet; siendo así que la Redención está tan entrañada en la expiación, porque una Hostia infinita expió una ofensa infinita; que la una puede tomarse por la otra, y decir que la Expiación o la divina redención es el centro de la historia de la humanidad, que es el principio de Bossuet; por lo cual se vé claro que el decir que este principio debe generalizarse aplicándolo a toda sociedad, no es audacia sino acatamiento al

Águila de Meaux: Si un sabio descubre un principio y le aplica a un caso grandioso, y un pobre escritor, a casos menos notables sólo con inquina nunca vista, puede llamarse como atrevido y pretencioso a éste y más si no niega que la expiación que Dios requiere en muchos crímenes, puede ser efecto de sólo su bondad, dada la virtud purificadora del dolor; si bien no diga que siempre, por no venirle ganas de negar el Infierno que existe a par de la Redención, aunque le tengan por comedor de perros y hombres crudos.

Nosotros, que siempre hemos tenido al señor Sañudo, al verle la casa, como una agua de malvas o que no ha inventado la pólvora, pasamos gustosos al Indio sus arrogancias, mas no cuando éste le acusa de no aplicar la ley de expiación a los hechos anteriores a Bolívar; pues decidnos Indio:

¿Tienes conciencia de que queriendo el señor Sañudo escribir algo sobre la biografía de ese, exijas se ponga primero a aplicar el principio a los suegros de Adán o tenga de irse a topar, con el maestro Caín, el envidioso, a decirle: "amigo, le cayó la expiación por la falta de respeto que t5 tuvo a don Abel cuando, sí mal no recordamos, le rompió las narices y le dijo hijue cabra que bien merecido se lo tenía su madre Eva, por sus conversaciones con comadres del barrio bajo.? No, achaque es de éstas, querer que se murmure de cuantos se puedan, aunque no vengan al caso, y principiari los cuentos desde donde el diablo perdió el poncho. Y luego ¿no te espanta ni asusta venir a citar un trozo en francés largo, sin traducirlo; y eso que ese trozo favorece más que empece al señor Sañudo, que escribió no sobre la historia de Colombia, sino sobre la biografía de Bolívar, para que se enojen los directores de Renacimiento? Porque si "la vida íntima de esos hombres (los notables) es alimento de la curiosidad biográfica" como dice el Indio y si "Les traits individuéis et vulgaires, qui le caractérisent, peuvent être releves pour les memoires et les biographies" como asevera el crítico de Cousin; rebien hizo el señor Sañudo al escribir la biografía de Bolívar y no la historia de un pueblo, de fustigar, v. g., a doña Manuela Sáenz, que públicamente y no a la tapada vivía con aquél en Bogotá.

Pero ahora sí que esos Directores que tan prestos berrinches tienen, se van a poner bravos por el trocito de francés que transcribimos sin traducirlo; A bien que no tenemos la culpa sino el Indio que le puso tan largo, tan mal traído a cuento y sin citar el autor, y eso que ya sabía, vayan viendo, el geniazo que se gastan esos señores. Por eso vamos a traducir nuestra partecita;.... pero no que el crítico de la Revista "Azul", nos ha de tirar las orejas; porque ha de decir que porque no inventamos el trozo, sino que lo copiamos, lo que no tenía gracia, de uno que estaba en francés. Ahí que se quede pues la traducción, aunque se enoje Apolo y todo el Olimpo.

Hay veces que los críticos no dejan lado a uno, como le pasó al pobre Sañudo, cuando dijo que el Señor había propuesto un término indefinido de progreso a la humanidad, con ciertas palabras que refiere la Biblia, de lo cual tomó pie el Indio, para haciendo derroche de erudición barata, decirle que había caído en un grave error; cuando habiendo dicho Sañudo que el término era indefinido, ya tácitamente declaraba que esas palabras se referían a una mera semejanza y no igualdad con Dios; pues lo indefinido dista de lo infinito infinitamente, de modo que podían parafrasearse de dos maneras con igual significado, diciendo, v. g.: Sed perfectos, o tan perfectos hasta asemejaros con el Padre Celestial; sin que el adverbio de la tina, haga errónea la frase; frase que como se refiere a la Caridad, a cumplir-se, haría la sociedad humana tan perfecta, como un paraíso. Lo más curioso es que el Indio atacó al señor Sañudo por aceptar el progreso indefinido, y el señor Velásquez Herrera por no aceptarlo, porque de veras no lo acepta. El Indio pues le

aderezó un error, para después darse el placer de devorarlo fácilmente, al igual que dice una antigua estrofa:

Tú te lo guisas,
tu te lo ten;
tu te lo comes,
Fray le mostén.

Hasta llegó a asegurar que había muerto sin esperanzas de resurrección; más el señor Sañudo debe tener el consuelo que desea su muerte el Indio, no por ser asesino, traidor o falsificador, sino por haber escrito una obra que tiende a honrar a Pasto, y a que, en lugar de que se levanten estatuas al venezolano Bolívar, se alcen más bien al caballero Nariño, o al inteligente Santander, excelsos colombianos, o con mayor acierto, al sabio Caldas, orgullo y honor de la ciencia colombiana. Su morir, pues, honraría la mediocridad de su vida.

Lo que motejamos al señor Sañudo, es su inoportunidad: Quiso el Gobierno hacer el favor a los nariñenses, de cambiar un fuerte de su moneda, con una peseta, y eso procediendo con tanta honradez para que no haya estafa, que les dio igual valor; con lo que se formó la moneda que decimos, con mucho garbo, de plata-oro, de forma que los alquimistas más famosos no nos llegan al tobillo; salió a oponerse el muy simple; siendo así que no teniendo la moneda antigua curso sino en Nariño, los valores tenían estabilidad, y los comerciantes, para hacer sus negocios, procuraban exportar frutos del país; con lo que impulsaban la agricultura e industrias nariñesas, dando a la par ocupación y holgura a los trabajadores; todo muy requeté malo, porque ahora, después del cambio de monedas, aunque los nariñeses no tengan trabajo, todo tienen los comerciantes con exportar la plata-oro, que hasta se está haciendo más rara que un tonto en el País. Otra vez, estando convencidos varios pastusos, que Bolívar en Bombona había dado a sus abuelos una paliza de Dios y mujer, y queriendo celebrarla con músicas y paseos, por ser muy donoso y de mucho gusto, festejar a quien vapuleó a los que les dio esta picara vida; vuelta a oponerse, alegando el muy sandío que los pastusos fueron los vencedores, como si para echar una cana al aire, mera menester averiguar bien quien venció a quién⁶³. Y se celebró el Tratado con el Ecuador, en el que se le cedieron los territorios de Sucumbíos, que siempre rigieron las autoridades de Colombia, y a oponerse Sañudo, siendo así que como dijo el señor General Carlos Guerrero; en el Congreso de 1924, el doctor Rodríguez Piñeres le refutó por completo, añadiendo el mismo General, que él tenía buenas razones para probar, que salió perdido por el Tratado el Ecuador; razones que manifestó no las exponía, por no ser secreta la sesión en que hablaba. Este motivo no lo entendemos en verdad, porque parécenos que era de exponerlas en la sesión pública, para honor y lustre de nuestra diplomacia; por lo que hasta creemos y apostaríamos que el señor General no tiene ningunas razones, bien que él dirá según un dicho, que ahí está jagracia del que es barbero, sacar patilla donde no hay pelo, pero cosa digna de espantar! en el número 258 del "Renacimiento", se vino a decir que varios publicistas han declarado que Colombia fue la perdidosa por el Tratado; por donde se ve, que no hay cómo tener renacimiento, y que con razón puso como empresa de su escudo, Mazarino: "El tiempo y yo contra todos". Por último: hubo época aquí que se quería más el ferrocarril del Pacífico que el de Tumaco, y se deseaba la denuncia

⁶³ Creemos que la mayor parte de los pastusos, celebraron la batalla de Bombona, sólo por su centenario, y no por otros motivos.

del Tratado comercial con el Ecuador; y Sañudo a oponerse a ésta, y a sostener que era principal para Nariño, el ferrocarril de Tumaco, por lo cual, respondió flojamente a “La Patria”, según afirma, con tanta sagacidad el señor Velásquez Herrera; y éste se desengañó.

Para acabar manifestamos a nuestros lectores, que aconsejamos al señor Sañudo, que no tuviera en cuenta las críticas del señor Ortiz, porque responder a ellas es como dicen los italianos: “trattare l'ombra como cosa salda” que es decir , tomar la sombra como cosa sólida, o sea aera verberare, porque aún este señor no le refuta, como dicen las personas sensatas, aunque sí está obligado a hacerlo, y punto por punto, porque lo prometió públicamente. Ojalá continúe el señor Ortiz su refutación que es un regalo, y si por ventura no le dieran hospedaje en los periódicos, deberá publicar hojas volantes, porque en todo hay que ser formal y de palabra y cumplir lo prometido. Y manifestamos también que, pues se ha hecho moda poner al pie de cada escrito la palabra continuará, también vamos a ponerla ahora, a menos que cedamos la mano a los amigos señores Pacífico Toro y Pascual Cordero, que quieren romper igualmente una lanza en el asunto.

¡Ah! se nos olvidaba contar algo de la crónica local: El 1° del mes de noviembre, se posesionó el nuevo Consejo Municipal, jurando cumplir las Leyes, y en seguida la mayoría dio una resolución contra la obra del señor Sañudo, que está consagrada a la glorificación de Pasto; protesta que no parece estar de acuerdo con el ordinal 7° del artículo 171 del Código Político y Municipal. ¡Al primer tapón zurrapa! También, aparecieron otras protestas, porque es muy razonable que a la cola de cada suceso se ate la correspondiente, y varios escritos contra el mismo señor Sañudo, de que, según datos, son autores o cooperadores los señores doctor Olegario Medina, General Carlos Guerrero, Roberto Hinestrosa, doctor Leopoldo López Alvarez, doctor Jorge Delgado Gutiérrez y Presbítero Teófilo Albán.

Sumáronse a los protestantes, los ilustrados Concejales de Mocoa, celosos guardianes del territorio y del honor de la República en el oriente, como dice el indio; porque efectivamente, por ahí, por donde esos valientes iba a atacar a Colombia, el libro del señor Sañudo. Parece, empero, que les hizo la protesta algún burlador para poner en ridículo la de la mayor del consejo de esta ciudad. Habráse visto mayor maldito!

Próspero Gallo

Unas Verdades (1)

Un veinte de julio histórico

El escritor bumangués señor Enrique Otero D'Costa, en un artículo publicado en el "Boletín de Estudios Históricos" número 4º, hablando de la obra llamada "Corografía de Pasto", dice: "Indudablemente... el estudio , del doctor Alejandro Santander es en alto grado meritorio; ayudado de muy escasos elementos de información, desprovisto de muchas fuentes históricas importantes, se vio forzado a hacer un trabajo en gran parte intuitivo, el cual, pese a sus lunares, pone de relieve sus facultades de, buen historiador y agudo crítico"; lo que es muy discreto, pues el acertar en la historia, sin tener auxilio alguno, no es fácil; por lo que la Obra, aquella Corografía, debe ser tenida como muy meritoria, y sus errores muy disculpables por las circunstancias en que se escribió. Fue el doctor Santander una de las personas que he conocido mas nobles y generosas, Patriota muy fervoroso y solícito buscador de datos históricos referentes a jeste país; por donde sería justo que se consagrara a su memoria, algún .recuerdo que demostrara la gratitud de sus paisanos, y yo aprovecho la buena coyuntura de recordar a la generación actual, sus méritos indiscutibles. ¿Pero qué mucho que mis palabras no encuentren eco ninguno, cuando he visto con sumo dolor, que el nombre impuesto a una calle de la Ciudad, del general Miguel Villota, que fue un patriota colmado y dechado de todas las virtudes públicas, no tuvo cabida en la última nomenclatura; siendo así que en las ciudades más ilustres se procura con esa modesta designación, honrar a sus hijos que por algún lado, les dieron ilustre fama?

Líbreme el cielo de incurrir en la torpeza de equiparar la Obra del doctor Santander, con la que publicó en ese mismo tiempo, José Rafael Sañudo, con el título de Apuntes sobre la historia de Pasto; sí busco empero que el trasunto de aquellas palabras del señor Otero D'Costa, sirvan para excusar ⁶⁴ siquiera en parte, los errores en que incurrió Sañudo; pues debe considerarse que se halló en las mismas circunstancias, con igual carencia de trabajos antecedentes y documentos, y en un terreno, puedo decir con todas veras inexplorado, hasta el extremo que hoy mismo, nada, absolutamente nada, ningunas noticias se tienen de los hechos que se verificaron en el siglo XVIII en esta Ciudad, y muy poco, casi nada, se ha adelantado en el conocimiento de los anteriores siglos de la vida de la misma.⁶⁵

Sé que nadie le ayudó, ni siquiera con mostrarle un documento; cuando ahora no es lo mismo, ya que el inteligente dramaturgo e historiador R. P. Anstídes Gutiérrez V. publicó varios, así como el doctísimo señor Obispo Pueyo de Val y el ilustrado General doctor Gustavo Guerrero, que por ello han merecido bien de la Patria. Porque si el ilustre R. P. Elíseo Villota, le obsequió con algunas copias de actas del cabildo de Pasto de los dos primeros siglos de su fundación, y el distinguido señor Felipe Obando Idrobo con algunos papeles, en que constan cuentas de la carnicería a fines del XVIII, fue doce aftos ha, cuando hace muchos de publicada su obra. A la cual obra, a pesar de

⁶⁴ Dice Gabriela Mistral: "Hay derecho a la crítica, pero después de haber hecho con éxito todo lo que se critica.

⁶⁵ Sólo he visto publicados sobre ellos, algunos artículos del Reverendo Padre Gutiérrez V. y unos sabios trasuntos del Ilustrísimo señor Obispo, y aunque según unas frases de Cieza de León paréceme que este se refirió a la fundación de Pasto en Yacuanquer como probaré declararlo algún día. Dios mediante, la labor del Reverendo Padre es docta y aprovechada.

que la escribió adolescente; siendo así que la adolescencia es la edad de la presunción que solo se corrige con el revólver del tiempo que engendra en todo hombre uno como escepticismo y desconfianza de la razón, por frecuentes desengaños, díóla el modesto título de Apuntes, y manifestó en el prólogo, cuan difícil es componer la historia de una pequeña Ciudad. Lo cierto es que han pasado 33 y 30 años respectivamente que salieron a la luz la 1a y 2a parte de ella, y hasta ahora, habiendo más empeño en reunir documentos históricos, más tesón en conocerlos y afición a la lectura de historia patria y hasta historiadores que tienen en poco la labor de otros que estuvieron en pésimas condiciones, aguarda el país una historia de Pasto que realice los deseos que expresó Sañudo en ese mismo prólogo, diciendo: "Quiera el Cielo que estos Apuntes compuestos por una mano tan indiestra, sirvan de apoyo a personas inteligentes y laboriosas, para que formen una historia digna de mi Patria y de la posteridad, haciendo, de este modo, cumplir mis más fervorosas esperanzas y deseos". No tuvo siquiera el ligero aliento, que sucede al escritor cuando hay lectores de su obra, que lo que es estimarla es pensar en gollerías, ya por lo mal compuesta quizás, y ya porque opónese a ello el egoísmo del paisanaje. ¿Pero qué si hoy a pesar de la existencia de una Ordenanza departamental, la 40 de 1917, el ex-Gobemador General Elíseo Gómez Jurado y su Secretario de Gobierno, señor Sofonías Ríaseos que juraron cumplirla, cuando juraron cumplir los deberes de su cargo y por ende todas las Ordenanzas del Departamento, no temieron ser perjuros primero que publicar la Obra, sobre Filosofía del Derecho, que escribió Sañudo, con el fútil pretexto de que antes una comisión debe conceptuar sobre su utilidad para el Departamento, menospreciando los nobles calificativos que le dio la Asamblea de este año, y con el más fútil aún, que expuso ese Secretario en su informe a la de 1926, de que esa comisión conceptúa sobre la identidad de la Obra, como si Sañudo hubiera escrito dos y la Comisión conociera la primitiva? El señor doctor Ángel Martínez Segura, apreciable por varios títulos, todo por ser un sabio jurisconsulto y afamado maestro de la juventud nariñesa, escribió en su informe como Rector de la Universidad de Nariño del año pasado, en la página 11, las siguientes palabras que aquellos ex-empleados debieran, para su propia enseñanza, meditar: "La Ordenanza N° 10, ya citada, no estableció condición ninguna, y a los encargados de cumplirla no les es permitido someter su eficacia a otras, que las impuestas por el Legislador Departamental. Establecerlas sería violar la lealtad, con que debe observarse el régimen departamental". Y no les sirve de excusa la pravedad de la Obra; ya porque primero es el deber no incurrir en un delito; ya porque el que un autor haya dado a luz algunas Obras malas, no es indicio seguro de ser también aquella, y ya porque el nombrar una comisión probaba que ellos no tenían certeza de su maldad. ¿O es que el nombramiento de la comisión no era sino pantomima para cumplir con su deber legal, sí la sabían? ¿Y cómo la sabían si jamás Sañudo les mostró la Obra? ¿No será quizás que algún desafecto o prevaricación se deslizó al cometer el perjurio?..... ¡Son paisanos al fin! No sucede así en otros lugares, y no quiero aducir el ejemplo de Santander de España, que asignó una cantidad a Menéndez Pelayo porque hiciera ciertas investigaciones; ni él de Maracaibo que fijó una pensión a Udón Pérez, porque se trataba de un profundo sabio y de un altísimo poeta; ni siquiera él de la Asamblea de Antioquia o él de la Municipalidad de Zipaquirá, que fomentan la publicación de las respectivas historias de ese Departamento y Ciudad, con publicar a su costa, las de los señores Ramón Correa y Luis Orjuela; si empero hace notar la conducta de esos empleados. Dios se vale muchas veces de los más menesterosos y humildes para producir un efecto bueno, por donde es posible haya querido que sea la Obra de Sañudo, de alguna importancia; de modo que quizás sea visto por las generaciones posteriores, cuando se sosieguen las actuales pasiones, como maravilloso el preferir un crimen con faltar al juramento, primero que editar la obra de un paisano.

Pero el colmo de la desestima del paisano, el colmo del antipatriotismo, el colmo de la ilegalidad, fue el protestar la mayoría del Concejo Municipal de esta Ciudad de 1925, que no debía por estimación propia y decoro patrio, con su presidente a la cabeza, contra la obra de Sañudo de Estudios sobre la vida de Bolívar; y lo que es más extraño siendo parte de la fácil protesta, los señores Juan Bautista Ortiz del Corral, quien dijo no haberla leído, Víctor Miguel Ordóñez y Diógenes Rosero, a los que siempre he tenido como personas de gran talento y de imponderable discreción. Porque tres fines sobresalen en dicha Obra, que tienen interés para el patriotismo, conviene a saber; defender a Pasto de injustas acusaciones principalmente; provocar uno como 20 de julio en la historia de Colombia, en que predomina la figura afortunada del venezolano Bolívar con ofuscamiento de proceres colombianos, muy superiores aunque con menor fortuna, tales como: Narifio, Santander, Córdoba, el cartagenero Manuel Castillo Rada tan infortunado que sus actos honrados han servido para vituperio de su vida, y el valiente José Padilla, asesinado con sin par injusticia; y apenas vista crueldad; y por último, escribir con toda verdad la vida pública de Bolívar. No es honrosa para un pastuso y sobre todo para miembros del Concejo, cuya es la misión principal mirar por el buen nombre y la fama de su ciudad patria que representan, tal protesta, y ni siquiera de un colombiano; pues se debe considerar que siendo un hecho bien averiguado que Bolívar volvióse enemigo de Santander, los últimos cuatro años de su vida porque éste oponiase a su desafortada ambición, si se le exalta por extremo, si se dice v. g. que era patriota, abnegado o lleno de valor, etc., lógicamente en la mente de extraños y aún de los propios, se engendra la idea que su adversario tenía los vicios opuestos a esas virtudes; es decir que Santander era antipatriota y ambicioso, cuando de modo fácil es defendible su actuación pública, y antes es digna de aplauso, salvo el fusilamiento de los oficiales vencidos en Boyacá, entre los que se contaban seis americanos y el humanitario panameño Jiménez, que no puede defenderse en ningún caso. Tal como escriben la historia los mismos colombianos, bien concíbese que el difunto vasco francés Marius André y el venezolano Blanco Fombona, aunque opuestos en doctrinas, sostengan que Bolívar debió fusilarle, pues le hacen aparecer como un monstruo de maldad e ingratitud. El sentimiento de xenofobia es reprobable, y criador de un insufrible chauvinismo, si se funda en una errónea apreciación, y yo mismo amo a los venezolanos Bello, Barait, y otros modernos como Dávila, Tavera, Mata etc.; peor no puedo menos de aplaudir con ambas manos, la conducta de los argentinos que glorifican a su paisano, por nacido en Yapegú, San Martín, el modesto y honrado militar, a pesar de que Ingenieros le tiene al par que a Bolívar, como no de talla de héroe; y no puedo menos de batir palmas a los chilenos y uruguayos, cuando pregonan como héroes nacionales, a O'Higgins y Artigas principalmente, y duéleme harto que en Colombia se pospongan a Bolívar, Nariño y los demás citados que con el sabio Caldas a la cabeza, son superiores por todos lados a aquél procer venezolano. En igualdad de condiciones debe preferirse a un primate de la tierra.

Oí a Sañudo decir a un respetable y simpático jesuita, a poco de escrita su Obra, que el 30 por ciento de ella sería refutable, porque creía que talvez algunos de sus juicios serían falsos, o que se aducirían documentos auténticos que declarasen hechos contrarios a los por él sostenidos, o que de los suyos se demostrase la falsedad; más hoy puede decir que no le han refutado ni el 3 por ciento, a pesar de que han transcurrido cerca de tres años de que salió a la luz su Obra y que algunos escritores notables entendieron en ella. Ni siquiera han podido defender ese error de Bolívar ejecutado cuando unió a la actual Colombia, Venezuela; sin parar mientes que la segunda no había de procurar la independencia de la Madre Patria, para quedar sujeta a un régimen central, a la primera, cuando estaba separada de ella desde la Colonia y acostumbrada casi a regir

sus propios destinos, siendo Capitanía General con ligeros nexos con el Virreinato de Nueva Granada. En el día, sería monstruoso siquiera el pensar la unión de ambas naciones en un régimen central, con todo haber rápidos medios de comunicación, como los ferrocarriles, calzadas automoviliarias, telégrafos, teléfonos y aún la aeronavegación, y más a principios del siglo XIX, cuando para llegar una carta de Bogotá a Caracas cuarenta días eran necesarios.⁶⁶ Paréceme, al meditar sobre ésto, que bien hizo Nariño en presentar un proyecto de Constitución de ideas federales, y que bien hicieron los federalistas que entonces existían en Venezuela y Nueva Granada; que bien pensaron Santander y otros, en la Convención de Ocaña, en implantar ese sistema, y hasta llegó a justificar a Páez que en 1826 quería la separación de Venezuela, salvo la ocasión ignominiosa de ser empleado del régimen central. Muchos escritores estimables por varios visos, se han ocupado de la Obra, como el patriota general Martínez L. los doctores reía Vásquez, Velasco⁶⁷ Carvajal, Botero Saldarriaga, Hernández y el señor Heraclio Uribe Uribe, ilustre por su persona y también por ser hermano del general Rafael Uribe Uribe, gran orador, hombre de letras y eminente táctico, quienes han declarado sobre ella sus sentimientos personales, si bien no han aducido documento alguno. Esto sucede con lo escrito por el doctor Luis Nieto Caballero, que siendo tan sagaz, llega hasta comparar una frase de Sañudo que sufre el genio de la lengua, con una, dechado del culteranismo o del hipérbaton latino que Lope de Vega en rias, proponía como ejemplar. Y llevado de su afición a Bolívar sostiene, como lo haría Nietzsche al decir que el hombre superior está más allá de la moral, que podía tener varias mujeres, pues había libertado varias Repúblicas, como si la moral sufriera condiciones y no obligara a los hombres extraordinarios. ¡Proh pudor! Por lo demás no sé que Bolívar tuviera, según él opina, arrosos como Napoleón en el puente de Arcóla, o Nariño cuyo caballo de campaña fue muerto en el primero al noreste de esta Ciudad, o Morillo en La Puerta en que salió herido, y Uribe Uribe en Peralonso; siendo cierto que el ejemplar de los jefes es un estímulo para sus soldados. Así escribe Augusto Nicolás en sus Estudios sobre el cristianismo: "Nada es tan contagioso y persuasivo como el ejemplo. Todos los tratados de patriotismo imaginables no hubieran hecho sobre el pueblo antiguo lo que hizo el desinterés de Régulo, ni nunca hubo una arenga que tuviese tanto valor como la acción de Conde echando su bastón de mando en medio de las filas enemigas, siendo el primero en arrojarse entre ellas para recobrarlo".

Muchos paisanos también, haciendo los Zoilos, insultaron a Sañudo por ese libro, con mucha sinrazón y arrimados sólo en su antipatía. No les había hecho ninguna ofensa para ello; antes bien a las mayor parte les había regalado sus publicaciones, y casi todos eran sus discípulos a quienes los había tratado con harta cortesía, y a alguno de ellos había suministrado datos históricos, y a otro, servido de Presidente de la tesis, cuando quiso optar al título de abogado, al mismo que cuando tesorero de la Universidad, dejó de pagarle algunos dineros; de forma que corresponde a Sañudo decir, como el político español Práxedes Sagasta, cuando le anunciaron que cierto sujeto era su enemigo, que no recordaba haberle hecho ningún beneficio. Decididamente la enemistad de ellos procede solo de la publicación de sus obras. Exinde irae. Extrañaba empero que un

⁶⁶ El hecho tan sonado, por ser el primero en una Coloma española, de la proclamación de la Independencia en Venezuela, que fue el 5 de julio de 1811, llegó a Bogotá el 21 del siguiente agosto, según la Patria Boba, escrito de esa época.

⁶⁷ Las Obras que ha publicado el señor Donaldo Velasco, no solo honran su patria que le vio nacer, sino las letras colombianas.

paisano llamado José Garcés Bejarano, no escribiese contra él, de suerte que cuando lo hizo encontré muy lógica su conducta. Al menos tenía un motivo contra Sañudo aunque especioso; pues encájesele en su magín, que él era obstáculo para que una noble señorita a quien cortejaba y a cuyo obsequio publicó unos versos muy amartelados, no le aceptase, cuando el verdadero motivo es que es cuarterón, lo que trasciende hasta en la color, que le ha valido el remoquete de "el negro Garcés". Además fue discípulo de Sañudo en la Escuela de Ingeniería, y era mal calificado, porque pese a su edad, era enteramente inhábil para las matemáticas. De allí su juicio sobre la novela La Espiación de una madre, que será muy mala, pero no es imitación de Cumandá, como es obvio para cualquiera que no sea de su laya; de allí el llamar feo vicio el hablar de sí un antor, porque eso se quería que éste no se defendiese así mesen injustos los ataques, y no existiese la entretenida lectura de las Autobiografías, como la que está publicando v. g. el señor Luis Mora, notable literato, o de las Memorias como la del abanderado Espinosa; y de allí que diga que Sañudo procede de una baja familia. Sí fue de una oscura, porque era pobrísima, y su padre prefirió una honrosa pobreza, primero que rematar bienes que se llamaban de manos muertas, a pesar de las sugerencias de un amigo, y su madre se dedicó a la enseñanza para vivir honradamente. Por lo demás me ha mostrado una documentación que consta de 31 páginas, que ahora pone a disposición de cualquiera, creada en la Montaña de Santander, en papel sellado español de 1802, en que aparece hasta la partida de nacimiento de sus abuelos hasta el quinto, con el fin de probar la nobleza de la familia Sañudo-Lloreda, y he visto en el archivo de la Curia episcopal, bondadosamente franqueado por el ilustrísimo señor Obispo, la comprobación de las cualidades y patria del malagueño Juan Torres Villalba en 1672, quinto abuelo también suyo; y de su bisabuelo viscaíno Manuel Sarralde Martínez Pinillos, la vista del promotor fiscal del obispado de Quito, doctor Flórez, de 6 de febrero de 1805, que expresa que "ha manifestado la información producida su hidalguía y los servicios hechos a su majestad en calidad de soldado distinguido". En cuanto a su bisabuelo José Damián Segura y Mendieta, natural de Bermeo, bien conocidos son sus descendientes en varias ciudades del antiguo Cauca, y como tenía escudo de armas por todos sus cuatro apellidos, que ahora posee su bisnieto señor Pedro Pablo Santacruz⁶⁸. Ciertamente que Sañudo solo ha conferido conmigo estas cosas pues tiene la convicción firme, de que es más preciada la flor que nace de un impuro pantano, que la en un regalado jardín, y que es más honroso tener ascendientes viles más que sean criminales, un hombre probo y recto, que quedar heredado uno, de un apellido ilustre, si su conducta es desdorosa para sus antepasados. San Francisco de Borja, duque de Gandía fue bisnieto del Papa Alejandro VI y sobrino nieto de los incestuosos hermanos Lucrecia y César Borgia, modelo éste del Príncipe de Maquiavelo, y fue empero varón santísimo y hasta General de la Compañía de Jesús.

Patricio Leal y Franco

⁶⁸ Fue casado en Popayán, en primeras nupcias con María Josefa Grueso, hermana del Provisor y Poeta, José María Grueso de donde procede Sañudo; y en Pasto, en segundas, con Dolores Santacruz, hija del último alférez real Gabriel Santacruz y Caicedo.

Unas verdades (2)

El prócer no tiene vida privada

Como venía diciendo en el último escrito, Sañudo hizo una conjetura no despreciable cuando dijo que el río Juanambú, "tomaría su nombre de Juan Ampudia porque antiguamente era escrito Juan Ampud y luego Juan Ambur", usando de esa forma dubitativa; porque ciertamente Ampudia fue el primer capitán que, con su hueste, lo pasó, y porque muchos ríos en Colombia tomaron nombre de los conquistadores, como el río Téllez en Nariño, Coello, Suárez y Lebrija, en otros lugares de la misma. Escribe Jaime Arroyo, en su Historia de la gobernación de Popayán: "Uno de los grandes afluentes de la opuesta banda (del Magdalena) fue desde entonces llamado Saldaña, porque en él se ahogó uno de los soldados que así se apellidaba; y más adelante, en la margen derecha, se dio a otro río la denominación de Cabrera por haberlo descubierto el capitán Juan de Cabrera". Y considérese que se decía Juan Ampud, que Joan no es quichuismo. sino la traducción al castellano del nombre latino Joannes; y que es del fonetismo castellano el convertir la o en u, de suerte que los baturros de Salamanca, dicen Tendura como nuestros labriegos pronuncian el nombre de Teodoro; pues parece que es ley en los romances el predominio de la u, de modo que aquel pronuncia bien el italiano, como los romanos, que palabras como por ejemplo coloro sonó, ellos son, pronuncia poniendo la boca como para emitir el sonido de la u. De allí el aforismo fonético, que es buena la lengua toscana en boca romana.

En el romance rumano abundan los sonidos de la letra, v. g. en surisu, sonrisa, flugin, no, vulpi, zorro, capul, cabeza, vulcan, volcán⁶⁹; lo que hace creer que en latín vulgar, sermo vulgaris, que coexistía a vueltas del clásico, de donde dicen que se derivan los romances que no del bajo latín que adoptó en su liturgia la Iglesia, predominaba esa letra. Con el andar del tiempo, puede que, trastocada la p en b, que es frecuente en el quichua, como veremos en Chaguarbamba compuesta de pampa, en Riobamba y otras, resulte Juanambú, suprimidas las últimas letras; lo que le dio la fisionomía de palabra del último idioma. Ejemplo clásico de cómo se transforman las palabras es Santander que, según un sabio, viene de Sant-andreas.

Entre nosotros, por obra de ignorancia, Antón Moreno que así se llamaban ciertas tierras entre Popayán y Timbío, por su propietario, hoy se llaman vulgarmente, Alto Moreno.

Puede usted, el amable lector, no venir en la conjetura de Sañudo, que al fin es una mera probabilidad; pero por vida suya, no de ligero convenga en la del doctor López Alvarez, que en la página 115 del número 4° del Boletín de Estudios Históricos, propone por estas palabras: "Aun sospechamos que el vocablo Juanambú sea también quichua". A bien que cinco líneas después la convierte en certeza agregando: " Se ha dicho que viene de Juan de Ampudia; pero ya se vé que es palabra indígena", y la razón es porque: "Si contraemos el vocablo Jua (nam) bu, nos resulta: Juambu o Huambu, es decir canoa". Ciertamente con síncope de las letras am en Juanambú, nos resulta una palabra quichua, y, lo mismo si cualquiera palabra, con gracia y donaire, sincopamos. Así Guatemala es quichua digan lo que quisieren los etnólogos que aseguran que es del maya quiche y que significa lugar cubierto de árboles, porque clarito se ve que si quitamos de ella el tema (que a veces hasta conviene pues es muy raro) nos queda guala que es el nombre quichua de cierta ave.

⁶⁹ Rumania la antigua Dacia, fue poblada por una colonia de soldados romanos o latinos.

Yo hasta aquí había creído, cuitado de mí, que en los nombres ya puestos, el vulgo por aféresis y metátesis, o a veces apocopándolos o sincopándolos, variaba las palabras, pero jamás de los jamases, que un lugar se nombraba añadiendo unas letras a una palabra; pero viviendo largo se aprende mucho, que al fin la experiencia es madre de la ciencia.

Cierto que en quichua, como en todo idioma aglutinante, uniendo palabras, o añadiendo a una, una sílaba semántica, esto es que de sí tiene algún significado, se da uno distinto o un matiz a la radical; más no que eso sucediera agregando unas letras, partes de sílaba, sin significado alguno. Y había creído, pecador que yo soy, que la toponimia muy repetida y más corroborada con otros datos, era base para la conjetura de que en los lugares hubo comunidad de idioma, y así sacar que chambú que hay en Nariño y cerca de Loja, y Mallama nombre de lugares de este Departamento y cerca de la Paz en Solivia, indicaban que los incas se extendieron a este territorio, y por no tener más datos, no concluía diciendo que Chile, cuya palabra es parecida a Chiles y que posee unos baños, llamados Colima, nombre de un volcán y ciudad de Méjico, tiene su población autóctona de la misma raza de ese Departamento o de esa Nación, ni que Juanambú sea Juan-varón, porque en Coaiquer varón se dice ambir, pero ahora ya sé gracias al doctor López Alvarez, que los quichuas se extendieron "hasta Neira, en el Departamento de Caldas y quizás hasta las montañas de Antioquia", porque hay allí un lugar como hubo en el Ecuador llamado Chaguallo, y "Guacaica, río que corre al norte de Manizales cerca de Neira, viene de huacuna, llorar, o huacai, lloro, y podría traducirse: he allí el que llora"; y hay en el Valle un lugar llamado Yanaconas y Tambo en el Cauca, con mal año para los historiadores, que contaron que del Ecuador llevaron los conquistadores muchos yanaconas, (que eran sirvientes en general, no de Atahualpa como advierte López Alvarez), que según Olano y Arroyo, su nombre significa cuello-negro; y construyeron tambos o posadas; y finalmente que también los quichuas se extendieron al Huila, por la fortísima razón que, "Uyumbe, colina situada en la región de San Agustín, Departamento del Huila, se deriva quizás de yumbuna, danzar, o yumbu, danzante".

Según la útil regla de él ¿hay más de sacar bonitamente un nombre cualquiera, que intercalándole unas letras o suprimiéndole otras, hasta llegar al que se antoja? Así es fácil decir y no hay trabajo ninguno para la etimología, que Lérida viene de Bogotá, y no se maraville el ingenuo lector de que una palabra esdrújula y europea, salga de una aguda y americana; pues recortándolas convenientemente se llega a la a que ambas tienen, y cada uno estornuda como Dios lo ayuda, y cada maestrillo tiene su librillo.

Hablando en serio, contaré que don Jaime Arroyo refiriendo precisamente el descubrimiento de Nariño, dice en la página 107 de su Obra: "Así tenemos que acudir en esta parte de nuestra relación a inferencias y conjeturas que deducimos, ya de los pocos hechos que constan como ciertos, ya del carácter de Benalcázar y ya, en fin, de la naturaleza del país y del estado social de sus moradores"; por donde se vé cuan razonable es hacer conjeturas, y también qué de excusable es caer en errores y más tratándose de la Historia de este Departamento. Recuerdo que el ilustrado catalán señor Alfredo Opisso que ha poco, dio a luz una historia de España y de las Repúblicas latinoamericanas, cayó en muchos, tales como decir que Nariño murió en 1816 en un castillo de Cádiz, cuan lo es cierto que finó en Leiva en diciembre de 1823; que Nueva Granada declaró su independendencia en 1915, habiéndolo hecho dos años atrás, en 1813 al 16 de julio; y que el venezolano Manuel Antonio Briceño, era natural de Cartagena de Indias. "El erudito Máñer denunciaba, dice el propio Opisso, 988 errores en los escritos de Feijoo, y Amiento y Soto añadían

algunas docenas más": ¿qué mucho pues si yerran los que escriben en lugares donde hay muchos medios de asertar, como ordenados archivos, caudalosas bibliotecas, etc. y tratándose de puntos en que ya hay abundante literatura, qué mucho de admirar si se extravían digo, en una Obra como Apuntes de Sañudo, sobre la historia de Pasto que es terreno inexplorado? Lo digno de estupor es que despotriquen, como si fuere menester lo demostraré, quienes tienen hoy más medios que no Sañudo, para asertar.

Más que todo sus Apuntes adolecen de vacíos, pues nada cuentan de los posteriores hijos de la viuda de Hernando de Ahumada, y de la familia de Díaz Sánchez de Narváez; nada de la poetisa que aquí vivió y murió y de sus hijos, cantada por Lope de Vega, en su Laurel de Apolo y recordada por Menéndez Pelayo, y de la ascendencia del Padre Lucero, sin otros datos curiosos de saberse, referentes a los siglos a que los Apuntes se extienden, además de las noticias que publican. Debe advertirse empero, que el error que les atribuye el Reverendo Padre Aristides Gutiérrez, cuando en la página 112 del Boletín de Estudios Históricos, asevera que Sañudo dice que Hernando de Cepeda casó el 1° de junio de 1562 con Gregoria de Zúfliga hija del Factor; no existe; siendo así que Sañudo al hablar en la página 36 del primer cuaderno de los Apuntes, de Doña Gregoria, se refiere a la esposa del capitán Ahumada desde el 16 de julio de ese año; pues el capitán Cepeda fue casado con Catalina hija del Adelantado Benalcázar, así como los capitanes Juan Cabrera y Pedro Velasco; porque si el Benalcázar no lo fue, tuvo varios hijos en mujeres principales incas, que fueron legitimados por rescripto del Príncipe, en cédulas de 1540, y Sañudo hasta da señas de un nieto suyo, Sebastián primer alférez real perpetuo de ésta ciudad, desde el 30 de mayo de 1594, por compra del empleo en 900 patacones. ¡Si los doctos como el Reverendo Padre yerran al atribuirle ciertos datos, qué no harán los que de Zoilos les vienen pintados la índole y el natural! Sin embargo hay en los Apuntes varios errores: de ellos, derivados de la tipografía, pues es de convencemos que aún al día de hoy, estamos en ese arte atrasados, y mucho más ahora 33 años; sin que esto reste mérito al señor Elias A. Villareal que los publicó, porque debemos considerar los elementos que tenía: un pequeño tórculo por imprenta y la oscuridad de la letra de Sañudo que hacía los yerros inevitables, de modo que antes de vituperio debe ser loado y tenido como benemérito; de ellos, por haber seguido, sin ninguna crítica, al Padre Juan de Velasco en varios asertos, puesto que ya conocía su blanda condición para creer inexactitudes, y que escribió en el destierro de Italia, más ganoso de hacer novedad en los extranjeros que de la exactitud de sus noticias, y cuya obra ha merecido muchas rectificaciones; de ellos, por ligereza en su proceder. Dado que no estoy de acuerdo con Gabriela Mistral, ni aún sabiendo que se refirió solo a los criticados que habían hecho una obra literaria, voy a dar mi opinión sobre esos errores. Entre los primeros está el que corre al ponerse que Sánchez de Narváez casó en 1583, fecha de su muerte, debiendo decirse en 1563, y al haberse llamado al capitán Juan Galíndez, que se le llama bien algunas veces, en cierta ocasión. Diego.

Entre los segundos, el haber puesto como fundador de Mocoa a Francisco Pérez de Quesada, que murió en Santa Marta en 1545, y de Barbacoas a Francisco Parada, que no se halla mencionado en esa época, sino el capitán Francisco Prado, que casó en Pasto, con Elena Cano, y aquí murió como varios de sus hijos, aunque sí fue teniente gobernador de Barbacoas, donde era feudatario por encomendero; y finalmente a Benalcázar como poblador de Madrigal, que solo nombró su cabildo. Entre los propios suyos, son el haber hecho a Cieza de León compañero de Benalcázar en la conquista de estos lugares, que por entonces estuvo en San Sebastián de Tierra-firme, y Sañudo le confundió con Francisco de Cieza, camarada de Ampudia; y él que campea de haber puesto al

primer Obispo de Quito, el apellido de Sánchez siendo Díaz; error que le advirtió el sapientísimo señor Pueyo de Val, y que lo tuvo quizá de que el nombre de Gara Sánchez es muy común, pues no solo fue de un rey de Navarra y de un lugar cerca de Arapiles en la provincia de Salamanca, sino de un poeta erótico del siglo XV llamado Badajoz aunque era de Ecija, y del metalurgista Carso que en el siguiente siglo, propuso la sustitución del azogue por el hierro, en beneficiar este último metal. Tal vez lo que influyó más en el yerro de Sañudo, fue que el primer cura de Popayán se llamó Garci Sánchez, que vino a estas partes con Ampudia, bendijo su bandera en 30 de noviembre de 1535, en la misa que celebró, según Juan de Castellanos, y tuvo por sucesor a Juan Pericón.

Aparece que la Audiencia de Quito, el 10 de junio de 1579, ordenó al Cabildo de Pasto que había admitido de cura a un Garci Sánchez, que quizás es el mismo que el anterior, que no lo admitiese y que lo hiciera salir de ciudad, que ni siquiera podía estar en las Indias; lo mismo a Miguel Gerónimo que había nombrado el Obispo de Quito, sin las formalidades del Patronato, si bien el Gerónimo no se creía cura sino vicario, nombramiento que pertenecía al Obispo. El Cabildo el 7 de octubre siguiente, dispuso que no se les pagase acostamiento ninguno. Tienen los Apuntes otros errores, como hacer a Benalcázar fundador de Pasto, y llamar al gobernador Tuesta del apellido Atiesta; más hago notar que en la Cronología de los gobernadores de Popayán, don Jaime Arroyo le nombra de la Puerta, y escribe que el doctor José Ignacio Peredo fue asesinado en Pasto en 1771, cuando fue teniente del gobernador Becaría Espinosa, que se posesionó el 23 de septiembre de 1777; pero en la culta Popayán no ha habido quien haya tratado de desacreditar esa Obra, a pesar de esos y otros errores.

Repúgname entrar a cuentas con Sergio Elias Ortiz; ya algo dije sin embargo de su consorte doctor López Alvarez, y es bien cierto que el lobo y la vulpeja ambos son de una conseja, y berzas y nabos para una son entrambos, de suerte que voy a contestar dos reparos que hizo a Sañudo. En 1588 se construyó el segundo puente que es al noreste de la ciudad, entonces entre la Carnicería y las casas de Lorenzo Martín, abuelo del jesuita Lucero, y de Bartolomé Chamorro, que tenía la suya en el mismo sitio de la del doctor Ezequiel Duarte; y Sañudo dijo en sus Apuntes que los vecinos dieron cuotas para su fábrica, y concluye diciendo que: "Así se reunieron 4.150 ladrillos y 179 pesos y así se hizo aquel puente etc.," en la página 55. Pues el señor Ortiz dice que en una lista solo aparecen 4-100 ladrillos y 100 pesos,⁷⁰ y que Chamorro únicamente dio 20, cuando Sañudo le supuso 30, en el número 2º del Boletín de Estudios Históricos, folio 36; al comienzo del 38 empezó, copia de una lista esos 30 de Chamorro, y en el 37 expone que después de la contribución: "faltó material y que los vecinos, sin que se sepa sus nombres, estuvieron listos a suministrarlo", y en la lista solo suma las cantidades que están expresas, que algunas no lo están, que dan los 100 pesos, y pone de contribuyente cuya cuota no está expresada, al capitán Díaz Sánchez de Narváez, que había muerto cinco años antes, cuando lo fue su hijo del mismo nombre; de modo que hace conocer que en el puente entraron materiales y costo más de lo que aparece en la lista expresamente.

También escribe en el folio 118, que el gobernador Mendoza y Silva no nombró a Diego Ordoñez de Lara, sino que confirmó el nombramiento presentado al Cabildo de Pasto en 2 de septiembre de 1606, expedido por don Juan de Borja⁷¹ para levantar recursos para la guerra de los pijaos,

⁷⁰ El cabildo también contribuyó con sus rentas.

⁷¹ Fue nieto de San Francisco de Borja

como lo decía: "cierto historiador por conjeturas"; pues fue el 27 de ese mes que al Cabildo presentó Ordoñez el nombramiento de Mendoza, en que se leen las siguientes palabras: "Don Vasco de Mendoza y Silva gobernador y capitán general desta ciudad de Popayán y su provincia.... mis lugartenientes, alcaldes ordinarios y otras justicias de la ciudad de San Joan de Pasto y Almaguer desta provincia, sabed que es a mi cargo en ella la conquista y pacificación de los indios pijaos.... y porque conviene que en las dichas ciudades de San Joan de Pasto é Almaguer.... se levanten y conduzcan todos los soldados efectivos de guerra que en ellas puede haber.... con sus armas e peltrechos y cantidad de indios amigos, é porque estoy ocupado en esta ciudad de Popayán y en las de Cali é Buga en levantar y conducir soldados.... acordé de cometer y encargar lo que se ha de hacer en las dichas ciudades de Pasto é Almaguer a Don Diego Ordoñez de Lara"; que si no es nombrar no sé qué pueda ser; cuanto más que en el mismo día y sesión dijeron los capitulares que obedecían: "la comisión del dicho gobernador y capitán general con la reverencia y el respeto debido"; que en lo transcrito no se habla siquiera de confirmar, y que confirma el superior no el inferior, como lo era don Vasco respecto de Borja que era "gobernador y capitán general en el nuevo reino de Granada, presidente de la real audiencia que allí reside", según se dice en la misma acta del 27. En la de 3 de octubre siguiente, también se habla de esa comisión. Los falsos asertos del señor Ortiz, tan falsos como los de Garcés Vejarano cuando aseguró que Sañudo vestía de un modo extraordinario, y en cierto tiempo, traje talar, y el afirmar el mismo Ortiz que en la novela Expiación de una madre, aparecen el cuervo de la de Isacc y el bambuco del Alférez real de Palacios, lo que es de todo punto falso, sirven para desacreditar las Obras de un paisano. No, jamás Sañudo ha cometido plagio, aunque sus producciones sean insignificantes: él jamás ha suscrito ni siquiera un cuento, que ha corrido ya publicado.

Patricio Leal y Franco

Unas Verdades (3)

La historia se apoya a veces en la conjetura

Quiero pasar en silencio, los dicterios que los Estudios de Sañudo dieron ocasión a que le dirigieran las protestas de la Sociedad "Antonio Ricaurte" y de varios artesanos, y otros escritos tanto de Colombia como del Ecuador, principalmente de Quito⁷² y de Pasto, reproducidos algunos en la prensa de esta Ciudad, y uno con introducción compuesta por el señor Roberto Hinestrosa y firmada por la redacción de la Revista Colombia, en que se comparaba desventajosamente esta Obra, y solo anotar cómo hubo quién propusiera a la Sociedad Bolivariana de Bogotá, la simpleza de que le declarara indigno colombiano, que por el Dios que me crió, no se me alcanza la facultad que tenga para ello.

Hay quienes encuentran buena la Obra en sí, pero juzgan que su publicación fue inoportuna, sin considerar que amén de que la verdad siempre es oportuna, si quería Sañudo sacar verdadero a Bolívar que en carta a Santander del 21 de octubre de 1825 decía que los pastusos debían ser aniquilados, porque sino recordarían a los cien años sus desastres, bien hizo al escribir y firmar la introducción en el centenario de esa carta.

Siempre fue el destino de sus Obras, aunque patrióticas, el atraer sobre sí el insulto y la diatriba. Recuértese que a poco de publicar un folleto en que probaba cuan perjudicial era para Colombia, el Tratado de 1916 con el Ecuador, no solo el reverendo Padre Prefecto del Putumayo, llamó ignorantes a los opuestos al Tratado, sino que el progresista ex-Presidente General Ospina, dirigió a los Directorios conservadores un telegrama, que los apellidaba de injustos y antipatriotas, y anunciaba que el Presidente de entonces doctor Concha, los había brillantemente refutado; telegrama que fue aquí publicado por dos caballeros de modo abundante. Y fue el caso que la refutación del doctor Concha, no era otra cosa que decir en substancia, que él respondía de lo hecho por un Canciller, como si viviéramos en un régimen parlamentario en que las Cámaras imponen los Ministros al Presidente, y no en uno presidencial, en que ese los escoge libremente, y por tanto, sin decirlo, es responsable de sus actos. Con justicia esa refutación no fue publicada en Nariño.

Hoy, algunos periódicos del Ecuador tachan de ingrata a Colombia, porque según su decir entre otros motivos, por ese Tratado la obsequió con miles de kilómetros de su territorio, sin compensación alguna; y no les falta razón si solo se tiene en cuenta que un Representante nariñés dijo en el Congreso colombiano de 1924, que el folleto de Sañudo había sido refutado; que él tenía razones por las que se veía que Colombia por el Tratado era favorecido; y que el Gobernador de esa época el señor Francisco Albán, que es de mucho talento en verdad, no solo dio un decreto en que calificaba el folleto con términos fuertes, sino que en la Imprenta del Departamento publicó las hojas que contra él escribieron, los señores doctor Eduardo Rodríguez Piñeres y Tomás

⁷² Ahora los quiteños tratan de alzar una estatua a Bolívar. Dios les valga por generosos! La Biblia aconseja, que no manda, que cuando alguno sea abofeteado en una mejilla, muestre la otra al ofensor, pero no que cuando se insulte a los antepasados de uno, se glorifique al insultador, porque cada uno es dueño de su honra, más no de la ajena, menos la de los ascendientes. Bolívar puso como chupa de dómine a los quiteños, en cartas de 1824 y 1823.

Márquez, y las hacía fijar en las esquinas de las calles por los empleados de la Policía, aplicando lo del Departamento a fines consiguientes que las Ordenanzas no quisieron; lo que repugnan los artículos 480 y 462 del Código Penal. Y lo más penoso para Sañudo debió ser, cuando sometida su conducta al juicio de la Asamblea, diputados tales como el malogrado poeta doctor Javier Santacruz y el doctor César Burbano, uno de los más inteligentes de Nariño, dieron a ella sus sufragios favorables.

Pero dejemos a un lado estos recuerdos, digamos únicamente que por lo dicho atrás, no puede deducirse que no debemos corregir los errores que se encuentran en los libros, sean de quienes fueren aún de paisanos; que la verdad a eso nos impele, que es más amiga que Platón, según dice un aforismo; pero guardémonos de toda malignidad y de echar a mala parte lo que no es evidente error; pues el error es con el hombre y no hay obra tan perfecta que no tenga alguno, ya que la naturaleza humana no es infalible; y por tanto debemos disculpar el que tenga evidente o probable disculpa, y el saber corregirlo sin el ensañamiento que solo cumple a los malignos. Así siguiendo está razonable regla, quiero advertir a Sañudo que si no encuentro en sus Estudios asertos que enormemente se aparten de la verdad, sí repeticiones de frases, y que paréceme que cuando narró que Obando derrotó a Mosquera en la Ladera con menores fuerzas, debió conjeturar menos, que sus tropas serían formadas de patianos, es decir del Tambo y de Timbío, que eran sus antiguos compañeros y ya veteranos, ando las de Mosquera serían milicianas, siendo averiguada, cuanto ceden aquellos a éstas en el arte de la guerra; y que el general Flores debió ser mediocre, como lo asegura en alguna parte, pues que durante más de veinte años dominó en el Ecuador. El ilustrado historiador señor Carlos Pereyra, mejicano que reside mucho tiempo hace en Madrid, también en una carta, le tacha de haber incurrido en un anacronismo, en la nota de la página 105, sin decir cuál fuese, seguramente porque en ella asegura que ciertas frases de una carta exponía Bolívar, para poder imitar a Iturbide, que se coronó como emperador de Méjico en mayo del siguiente año de la fecha de esa carta, de modo que no pudo ser imitador, aunque debe advertirse que mucho antes de coronación de Iturbide, se sonaba el monarquismo de muchos mejicanos. Igualmente en los Estudios aparece que el batallón Numancia pasó a los independientes en 1819, siendo cierto que eso hizo el 3 de diciembre de 1820, después de la insurrección de Riego. Sé positivamente, pues vi los originales que escribió Sañudo de los Estudios, y que muy muchacho leyó la Obra en que el filósofo de Ginebra, Dusseau, hace el maestro de un rapaz a quien llama Emilio, antes de saber que esa Obra, que lleva ese mismo nombre, estaba en el índice de libros prohibidos, que es mera errata lo que aparece en la página 2a ; los Estudios, al decir Emilio Rousseau, con supresión de la preposición de, y que también es, la que consta en la página 3a; pues estando en los originales que las Cortes de 1812, propusieron la solución de algunas dificultades a Bentham, al cambiar el 2 en 8, quedó un tamaño error; porque en 1818, jamás existieron Cortes, que Fernando VII, por Decreto de Valencia de mayo de 1814, las suprimió con todo el régimen representativo. Finalmente me consta que el tipógrafo cambió hecho la ortografía.

Tiene Sañudo disculpa en haber pasado en silencio los primeros años y estudios de Bolívar y su vida íntima; pues el ámbito de unos estudios puede restringirse a voluntad y hasta a una sola faz; de modo que por este viso, no merecen la nota de incompletos, cuanto más que hay almas tan delicadas que se ofenden con la más leve sospecha de inmoralidad; quisiera empero que Sañudo considerase no solo la vida pública e Bolívar como se propuso, sino también la privada, porque ésta sirve para conocer aquella, que la vida de familia es como un trasunto de la pública y sírvele

como preparación⁷³. El procer siempre es el mismo, con iguales pasiones, en la sociedad familiar que en la del Estado; y que no se escandalicen algunos, alegando que es intocable la primera o del hogar. Básteme en justificación aducir las palabras del eminente polígrafo argentino doctor Ernesto Quesada que en su Obra "La época de Rosas" escribe: "La historia, se ha dicho, tiene que hacer con los grandes lo que hace Dios con todos, y para eso tiene que ser imparcial y verídica. No sirva de disculpa lo que se llama el respeto a la vida privada: los reyes y los gobernantes no tienen tal derecho adulator; el que quiera que se respete su vida privada que se reduzca a vivir como particular, pues los que se ponen muy en alto son muy vistos y sus vicios enseñan a los pequeños a cometerlos", y en una carta expone; "En cuanto a que "la vida privada es territorio vedado a la crítica histórica, ¡hum! Ya no puede considerarse ese noli me tangere sino como un anacronismo, en las postrimerías de este siglo (escribía en 1898) y después que la novísima crítica histórica ha desmenuzado sin piedad la vida de los hombres que han influido en la marcha de los pueblos, para encontrar así el secreto de sus acciones o la explicación de los acontecimientos". En cuanto a los primeros estudios de Bolívar, es cierto que poco podía escribir Sañudo, porque un curador de aquel procer, mandó estudiar a la Península, cuando tenía diez y siete años y aún no sabía siquiera ortografía, y solo estuvo un año en el colegio de España, pues después, perdido de amor por su prima María Teresa Toro, optó por casarse y mientras obtenía la licencia, viajar por varias partes de Europa. Alguna vez dijo que tuvo muchos maestros, pero vayan ustedes a creer que eso prueba su aprendizaje, ¡Medrados quedarían!

Que deben tener errores los Estudios, a priori debe deducirse, pues como ya dije, muy humano es el error, pero no están en el reparo que hizo el Reverendo Padre Fray Heliodoro de Túquerres, en el primer artículo que salió suscrito con el seudónimo de El indio patriota; ya que los posteriores no son suyos, y no va nada que cualquiera quiera atribuírselos, porque son oscuros y laberínticos por demás, yo a lo menos no los entiendo. En efecto el principio de la expiación, a mi juicio, no solo domina en la Religión sino también en la Moral. Platón y San Agustín lo proponen en la Antigüedad, Santo Tomás y el Dante lo sustentan en la Edad Media; Maistre, Donoso Cortés, Augusto Nicolás, aún K-ant, los más grandes ingenios lo propugnan en la Moderna; y los dogmas de la Redención y del Purgatorio son expiación mezclada de amor y de esperanza, y del Infierno es expiación colmada de justa eternidad.

Antes de entrar a hacer algunos reparos a los Apuntes sobre la historia de Pasto, paréceme bien advertir que en la historia es de uso frecuente la conjetura, y si es acertada y si nuevos documentos demuestran su exactitud, sirve para lustre y loa del conjeturado; que así justamente por eso goza de gran renombre el historiador Michelet, cuyas conjeturas se han verificado. Lo pernicioso no es proponerlas, en la historia o en la ciencia, que por ellas o lo que es lo mismo, por las hipótesis, se llega a la verdad, que en parte es inasequible sin un laborioso discurrir, o no se le alcanza del todo en la mayor, como que es infinita; sino el ponerlas como una verdad asentada o dogmas indiscutibles, y a veces con liviano fundamento. Cuando son provechosas fúndanse en la probabilidad física o moral, lo que no es de despreciar en modo alguno; en prueba de esto es de considerar que aún estamos en Cosmogonía en las conjeturas de Kant y de Laplace, que les han dado gloriosa fama y nombre singular, y que ya Platón las tenía como fuente de conocimiento, como lo expuso en el libro 6° de su República. ¿Y qué es decir si la tradición local favorece la

⁷³ Dice Henry Joly en L 'avenir français: "Que ja moralité publique et la moralité privée, agissent et réagissent l'une sur l'autre c'est ce que personne en niera".

conjetura o le da los fundamentos en que se apoya? ¿Y qué entonces si algún historiador de acuerdo con esa tradición y la probabilidad moral sostiene, pongo el caso, que Nariño no se presentó ante el pueblo de Pasto, cuando preso en la casa del cabildo, con ese aire de espadachín con que le expresó en el bronce un escultor extranjero, que nada le iba ni venía con el carácter elevado del héroe, ni conocía su historia ni menos su grandeza? Suponer que no estaba preocupado por haber perdido una hermosa expedición en que entraron la flor y nata de jóvenes de varias Provincias, treinta y seis oficiales distinguidos y muchos soldados que eran el nervio de su juventud, después de haber obtenido los gloriosos triunfos de Palacé, Calibío y Tacines, que ganó, nótese bien, antes de la insurrección de Riego y Quiroga, que puede decirse causó la Independencia de la América española; superado los grandes obstáculos que entonces oponía el Juanambú, y malgrado los frutos de su campaña, dando ocasión a que le zahiriesen y dieran vaya! sus enemigos que nunca faltan al hombre de mérito; es desconocer el corazón humano; es criar una especie tonta de estoicismo que jamás existió; es ignorar el carácter pundonoroso de Nariño, que según Soledad Acosta de Samper, era sensible, nervioso e impresionable, y que en algún paso de su vida se mostró triste y pusilánime, y atribuirle uno cual la conducía de un cretino.⁷⁴ Provocar al pueblo de Pasto en esa ocasión tal como le representa la malhadada estatua que hay en la Plaza mayor, es gesto de un necio, que la resignación a la desgracia, que es de fuertes varones, no ha de confundirse con la insolencia. Yo por mi sé decir que voy en la conjetura más adelante, más que me apelliden como quieran, pues pienso que Nariño voluntariamente se hizo coger prisionero, que magnánimo y delicado no sufrió el volver a Bogotá con tan tremenda desgracia, porque sus émulos envidiosos le denostasen; ya que pudo ponerse en salvo, como muchos y aún su hijo lo hicieron.

El artículo 588 del Código judicial define la conjetura diciendo que es "el juicio favorable que formamos sobre la verdad del hecho controvertido, en virtud de datos falibles pero no despreciables"; y el artículo 541 propone como pruebas legales, después de la confesión la presunción legal, los indicios o conjeturas, antes del testimonio, escrituras e inspección ocular y fama pública; ¡o que prueba la importancia que les da.

Por el consiguiente tengo por poco cuerdo al que denostase a un escritor por sus conjeturas racionales. Vaya el ejemplo del notable historiador venezolano doctor Vicente Dávila, quien en la importante Obra, que, según el colofón de ella, se acabó de imprimir en el pasado último octubre, y que se denomina "Investigaciones históricas" al hacer la biografía del General Jacinto Lara, dice lo siguiente: "En su vida pública hay un acontecimiento que deslustra su brillante hoja de servicios: en mayo de 1817 la ejecución⁷⁵ innecesaria de los veinte Capuchinos civilizadores del Caroní, en cumplimiento de órdenes de Bolívar. Estas debieron de ser verbales, hasta hoy la historia no habla de ningún documento sobre el hecho. Pero que hubo la orden es también una verdad histórica, donde no Bolívar habría castigado semejante crimen cometido casi en su presencia. Que no era el Libertador hombre que soportara un atentado como ese, ni tampoco en lo militar ejecutan los Tenientes tales hechos, sin estar facultados para ello. Puesto que a no ser así la disciplina, que es el alma del ejército, no existiría".

⁷⁴ Hasta Jesús, Vida nuestra, a pesar de la infinita fortaleza de la Divinidad demostró con su sudor sanguíneo, la amargura que su Pasión le causaba, cuanto más un débil mortal.

⁷⁵ Caruche, añadido yo.

Tiene Sañudo en sus Apuntes una conjetura muy plausible, fundada en la fonética del castellano en buena parte; pero una vez que va bastante largo el actual escrito, mejor la dejaré para decirla en el siguiente, que haré si Dios me diere licencia para hacerlo.
Patricio Leal y Franco

Unas Verdades (4)

En Junín y Ayacucho, Pasto no se distinguió por el amor a la libertad

He creído que pues había corregido los errores que encontré en las Obras jañudo, debía con mayor razón, rectificar los que constan en el Boletín de Estudios Históricos, que plugo a la última Asamblea se publiquen por el Departamento, porque alguno no vaya a creer que todo lo expuesto en ellos es verdadero, y así se propaguen noticias falsas como verdades ciertas. En el presente escrito, voy por tanto, a demostrar algunos y a la vez contar hechos curiosos de saberse, apoyado en documentos fidedignos toda veracidad.

Sea lo primero que en el sobredicho Boletín, en el número 2 página 35, se : "En los primeros días de noviembre de 1587 llegó a esta ciudad, a practicar visita oficial el Gobernador de la Tuesta Salazar, acompañado de su Teniente General, el licenciado Joan de Velásquez" y en una nota que "En ninguna de las actas del Cabildo de esta ciudad aparece actuación alguna del Gobernador de la Tuesta Salazar, si no es hacia el mes de noviembre de 1587" y sin embargo en el acta del 15 de julio del mismo año, ya aparece en Pasto, porque señaló en ella, los regimientos que se debían dar a Doña Mana de Montoya para ayuda de costas del casamiento de su hija Doña Luisa, y además el número de regidores que debía tener el Cabildo de esta ciudad. Consta por la del 17 de octubre siguiente, que nombró de su lugarteniente general al licenciado Velásquez; de modo que no es cierto que en noviembre hubiera venido a la visita acompañado de su Teniente General, porque aquí fue nombrado Velásquez en lugar del capitán Francisco de Olivos, que se asiló en una Iglesia porque estaba acusado de graves delitos y que el 14 de abril presentó al Cabildo su título. El 27 de enero, del siguiente año, el Cabildo y el gobernador acordaron que se hiciera el puente que está al nordeste "entre la Carnicería y las casas de Bartolomé Chamorro y Lorenzo Martín". Este Martín fue el abuelo paterno del Padre Lucero, al que en la misma página el Boletín llama Muñoz, y en concordancia con el convenio, el acta del 1 ° de abril se lee: "y todos de un acuerdo y parecer trataron de que se encargue la solicitud dello a Pedro Gaspar y la distribución de lo que en la obra de la dicha puente se ha de gastar y que el dicho Gobernador dé mandamiento con que los propios desta ciudad entren en poder del Pedro Gaspar hasta entanto que la dicha puente se acabe y se le dé libramiento de las penas de cámara y de todos los propios desta ciudad que estuvieren hasta aquí a su cargo, y en los demás no se pueda librar cosa alguna fuera del salario que se le da al presente escribano hasta acabar la dicha puente". En el Boletín se copia que ella se construyó: "a costillas de la Ciudad y de las personas que mandaren algo", añade empero que "Al decir la ciudad, no se crea que fuese el Cabildo el que ayudase en los gastos; no, la ciudad y las personas, eran la misma cosa. Ni al Tesoro Real o Real Cámara, que decían entonces, ni al en aquellos tiempos paupérrimo Tesoro Municipal, les mereció la obra un solo maravedí. Fueron los vecinos, y solo los vecinos los que con sus contribuciones levantaron el mencionado puente, etc.". Con razón Sañudo dijo en sus Apuntes, no fijándose solo en la cuestación de los vecinos, que "Así se reunieron 4150 ladrillos y 179 pesos, y así se hizo aquel puente", que jamás exceptuó la contribución de la Ciudad para su fábrica. El Capitán Díaz Sánchez de Narváez, el que ganó escudo de armas por sus valentías en Jaquijaguana, no contribuyó porque era muerto, cinco años antes, sino su hijo del mismo nombre, Zúñiga por su madre.

En el número 1° página 13, los directores del Boletín dicen: "Cuando en 1809 el pueblo de Quito dio gritos de libertad, muy humano era que nuestros pueblos rechazasen esos primeros brotes,

cuando el resto del continente permanecía callado y en expectativa"; lo que es inexacto, porque antes de que Quito formara la Junta que reconoció la Central de España y la Regencia, ya la había creado Montevideo, y en 25 de mayo de ese año, Chuquisaca, y en 16 de julio La Paz, que proclamó abiertamente la Independencia. En la página siguiente dicen: "Pero la época más brillante del patriotismo y del amor a la libertad de los pastusos es la de los épicos días de Junín y Ayacucho.

Algo más de mil hombres... fueron sacados de la ciudad y llevados a combatir al Perú"; pero olvidan decir que fueron reclinados traidoramente que fueron al Perú, mal de su grado. La verdad ante todo.

En la misma página 13 se dice que hubo: "la sublevación de los pastusos contra el Gobernador de Popayán, don José María Paredo, que culminó en la muerte de éste a manos del pueblo de Calambuco, el 7 de noviembre de 1810", lo que no es nada cierto, pues no hubo un gobernador de ese nombre en Popayán; lo que aconteció es que Piñeres, visitador de Nueva Granada, impuso un estanco de aguardiente, y el gobernador Becaria Espinosa, que se posesionó de ese destino el 23 de septiembre de 1777, envió a su teniente general doctor José Ignacio Peredo, que además era administrador de los bienes de los recién extinguidos jesuitas en la gobernación, a establecerlo en Pasto, y aquí, del colegio de la Compañía, fue sustraído por un tumulto popular y muerto en Calambuco, por un tal Naspirán, un jueves 24 de junio de 1781, como más largamente refiere Sañudo en sus Apuntes que están inéditos. Ya en marzo de este año, había estallado por el mismo estanco la sublevación de los Comuneros y después los motines de Mérida en Venezuela, Barbacoas y Tumaco; pero el Virrey concedió una amnistía, de modo que cesaron los procesos contra los fautores. La muerte de don Francisco Clavijo, que en la misma página mencionan los redactores del Boletín, ocurrió en mayo de 1800, con la de su hermano paterno el joven Atanasio, por el recudimiento de los diezmos que se había decretado, y no por ideas de Independencia.

En el número 3, aparece un discurso del señor Ortiz y en la página 66 se dice: "porque los padres concriptos de ese Ayuntamiento, (el de Pasto cuando Nariño le dirigió notas que fueron rechazadas) eran hidalgos de la más dura corteza, un Arámburo, un Tinajero, un Ortiz, un de la Villota, un Zambrano, etc." cuando revisadas las contestaciones del Cabildo de 1° de abril, 4 y 8 del mismo mes, no figura en el Ayuntamiento ningún Arámburo ni Ortiz; si bien fue miembro un Ortiz en años anteriores, de que es pariente el actual alcalde señor Juan Bautista Ortiz, uno de los más caballerosos funcionarios. Lo componían los señores Ramón Bucheli, Mariano Santacruz, Ramón Tinajero, Francisco y Manuel de la Villota, Francisco Hidalgo, Blas Rosero, Francisco Zambrano y Eusebio Burbano.

En el número 4° página 118, dice el señor Ortiz que Diego Ordóñez de Lara: "murió en 1579, ejerciendo la gobernación". Más don Jaime Arroyo, en su Historia General de Popayán, escribe en la página 300, que se posesionó en Cartago el 11 de diciembre de 1589, y en la siguiente, que murió el 26 de enero de 1591; y en las actas del Cabildo consta que el 23 de diciembre de 1589 el Capitán Hernando Cepeda Caraveo, presentó al Cabildo su título de gobernador; y que a este Cepeda el 1° de febrero siguiente, nombró de su lugarteniente, y que el 23 del mismo, el licenciado Cristóbal Bótanos presentó el título de subteniente general, que dice ignorar quién fuese, el señor Arroyo. Cepeda dejó la vara de su cargo, el 9 de febrero de 1591, porque expresó que sabía que Ordóñez era muerto. En 3 de octubre de 1606, los cabildantes apelaron del mandamiento del

visitador Armenteros y Henao "y de la comisión librada por don Vasco de Mendoza y Silva"; porque es de saberse que este gobernador de Popayán nombró al otro Ordóñez de Lara, para que en su lugar, levantara tropas en Pasto y le exigiera recursos, para la guerra de los pijaos. La Audiencia de Quito por provisión del 20 de noviembre, y no del 28 como dice el señor Ortiz, ordenó, que la Ciudad diese 15 soldados y 100 pesos para dicha guerra, lo que se notificó a Ordóñez de Lara el 1 ° de diciembre, que dijo en cuanto a su cumplimiento: "que obedece lo que su magestad manda por agora, con protestación que hace de *informar* y pedir a su magestad y a *la dicha real audiencia* más bastante socorro y que esta ciudad lo dé para la dicha guerra..."⁷⁶

El Ordóñez dio un auto, de que se notificó el Ayuntamiento el 4, en que disponía que el Cabildo, entre otras cosas, se obligase a dar los 100 pesos y los 15 soldados con su avío, "el cual avío ha de ser en cuanto a las armas diez arcabuces y cinco picas, y para cada arcabucero cuatro libras de pólvora y cuatro de plomo y tres de mecha, y para cada soldado treinta pares de alpargates y dos mantas para un escaupil (sayo) y algodón para hacerlo, y para su bastimento y sustento de cada soldado doce arrobas de bizcocho y seis jamones y cuatro arrobas de sal..." El mismo día el Cabildo manifestó que estaba pronto a cumplir lo mandado por la Audiencia, "y en lo demás tocante al excesivo precio que el dicho capitán pide en el avío de los soldados...responden que atento a que en la real provisión no se declara más de que dando esta dicha ciudad 15 soldados y 100 pesos de 20 quilates para su avío hayan cumplido, y para que haya cumplido efecto lo que a cargo de este cabildo está se trató de dar para cada soldado 100 pesos de oro en polvo de Almaguer y los 100 de veinte quilates, que todo viene a ser y monta un mil y setecientos pesos con que cómodamente se pueden aviar los dichos soldados..." Después en el Cabildo de 11 de diciembre, a esa cantidad aumentaron 300 pesos más y 100 arrobas de bizcocho, y se comisionó al capitán Sebastián de Benalcázar, al alcalde Martín Sáenz de Ojeda y al regidor perpetuo Joan Guerrero no Villafañe, para que hagan el reparto entre los vecinos de esas cantidades. Véase pues que el señor Ortiz confunde las exigencias del auto de Ordóñez con lo que se obligaba el Cabildo; como a pesar de que transcribe las frases de una historia que dice: "plagada de errores" de que Ordóñez por los 100 pesos exigió los artículos que en ellas se indican, cree que esos solo fue lo que dice esa Historia que llevó, cuando ésta no declaró únicamente las minucias del asunto.

En el número 5° y en la página 153 dice el señor Ortiz: "Como se ve por el memorial anterior y el decreto, en el tiempo a que se refiere el autor de los apuntes, había más de una venta hacia los lados de Juanambú". Pues no se ve leídos atentamente, a no ser que tambo o posada y venta se tomen como homónimos. En el mismo y en la página 154 dice el Dr. López Alvarez: "...a principios de nuestra era existía ya organizado el primer imperio peruano". Quien sabe, algunos eruditos lo hacen posterior al siglo X. En el 6°, página 168, pone el señor Ortiz una lista de las personas que contribuyeron para el reloj de San Juan Bautista y antes expresa: "lo hacemos por dos objetos: para dar a conocer quiénes componían la buena sociedad hace 80 años y para dejar una especie de constancia de la gratitud de la ciudad para con ellos". Lo segundo es de aplaudir, pero en lo primero debe verse que no es una guía segura un donativo, para conocer una buena sociedad. Los señores que están en la lista, fueron notables por varias acciones, y no solo por contribuir a la compra del reloj, y por otro lado, el motivo de una idea falsa, por incompleto, de la buena sociedad de aquel tiempo. Muchos que no contribuyeron con sus dádivas, quedarían excluidos justamente creería alguno, cuando hubo varios notables que no están en la lista, que recuerdo, sin

⁷⁶Lo que está en bastardilla es lo trastocado o suprimido en la transcripción del señor Ortiz.

que pretenda agotar el número. En efecto, no consta el nombre del señor Fernando Hinestrosa, cuyos hijos son ornato de nuestra sociedad, y que formó una noble familia; ni el señor Joaquín Santander, de quien basta para conocer el lustre, su hijo el señor Modesto Santander, 10 de cuyos descendientes es hoy inteligente fiscal de un Juzgado Superior. Ni aparece el nombre del coronel Manuel Barreda, hijo del español don Juan, que fue gobernador de la Provincia de Pasto. El coronel Barreda tiene el honor de ser ascendiente del honrado y nobilísimo administrador señor Wenceslao Gallardo. Ni el del valeroso y humanitario coronel Custodio Rivera, cuyos descendientes como Dídot para la Francia, son excelentes tipógrafos, ni el coronel José Folleco, el coronel José Félix España, padre del inteligente patriota y abogado doctor Aristides España, don Pedro de la Rosa, que lo fue del general Pedro Marcos, vencedor en Silvia de don Julio Arboleda, los doctores Manuel Córdoba, Agustín y Manuel Guerrero, los señores Rafael Guzmán y Juan Arturo, los Padres Manuel Guzmán, Gabriel Rojas y Fajardos, uno de cuyos parientes es el señor Luis Fajardo, que escribió un cuento que es honor de las letras de Nariño; el profesor José María Rojas Polo; el doctor José A. Chaves, único pastuso que fue ministro de Estado y que cuenta entre sus descendientes al doctor Federico Puertas, de los mejores escritores de Colombia; el doctor Maximiliano Chaves, que tiene, que tiene el honor de ser padre y abuelo respectivamente, de los notables abogados don Elías Chaves y doctor Bolívar Chaves; y el doctor Lucas Vergara padre de tres buenos sacerdotes y del distinguido Magistrado doctor Luis Vergara. Además los señores Juan Rodríguez y coronel Francisco Uscátegui y el P. Francisco de la Villota, fundador del Oratorio.⁷⁷

En el número 7 página 208, hablando de las comisiones para el encerramiento del Santísimo Sacramento el Jueves Santo, se dice: "Por el contexto se ve que era costumbre antes de 1610, pero no hemos podido encontrar constancia alguna anterior"; en el Cabildo empero del 30 de abril de 1584, nombraron los capitulares al fiel ejecutor para ello en uno de los monasterios, y agregan: "y esta horden se guarde de aquí en adelante". Ordenó Carlos V que cuando se fundara una población, se fundase también un hospital, y así en abril de 1562, estaban en Pasto trabajándolo con el nombre de Hospital de la Madre Dios, como consta del testamento de Miguel de la Barría, hijo de Joan de Sanabria y María Loyola, naturales de Oyarsum, que dejó una manda para él. Era por el consiguiente real y no eclesiástico; de modo que el Cabildo nombraba desde 1564 su patrón que lo fue el capitán Joan Crespo; y si el obispo de Quito fray Pedro de la Peña, lo confió en 1572 a los dominicos, hubo de cejar, porque el Cabildo siguió nombrando su síndico, y fue de suerte que pasando para el Perú, el Padre hospitalario Francisco de Villegas, le hizo propuesta de que los Hermanos de San Juan de Dios se fundasen en él y pusieran botica, con tal que se les cediese su renta; a lo que los capitulares en la sesión de 11 de agosto de 1671 "Unánimes y conformes nentine discrepante" vinieron con ello. No se sabe en qué paró la cosa, y de paso anotaré que el Obispo tenía alguna ingerencia en los hospitales por razón del culto que allí se daba. Duró el de aquí hasta que el Obispo de Quito, Alvarez, el 11 de junio de 1797, declaró el lugar profano, y se pensó de acuerdo con el gobernador de Popayán Diego Antonio Nieto, pasarlo a un lugar del barrio de San Andrés, porque su capilla era caída; de suerte que se ordenó borrar todas las señales de que era santo el lugar, al vicario Manuel de la Barrera, que asignó el 26 de febrero siguiente para su cumplimiento. Desde entonces acabóse el hospital, del que no ha quedado más recuerdo, que su nombre en la calle adyacente; pues tenía su asiento y una corta plazuela, en lo que es hoy casa de la familia del caballero y honrado señor Nabor Enríquez. En los que allí

⁷⁷ No he puesto todos los descendientes de los notables citados, porque sería alargar mucho el artículo, que solo he escogido los más benévolos para Sañudo.

murieron, cuéntase el Padre Antonio Ribeiro, de Gagos en Portugal, que tullido desde 1687 murió cuatro años después. Por lo narrado se vé que es impropable lo que asegura el visitador don Luis de Quiñones, de que un Rosero fue el fundador de él y la conjetura del señor Ortiz que dice en la página 231 del Boletín No 8, que el documento que cita "nos da a entender que su institución se debió a iniciativa particular", y también lo de la página 233 que termina así: "Con esta posesión, quedó establecido el convento que más adelante se llamó de Santo Toribío y que subsistió... hasta principios del 1840, en que fue suprimido, etc."; pues a más de que el testamento de Jorge Toro, natural de Lorca, hecho en 1564 donó cuatro pesos para el dicho convento "que se está haciendo en esta ciudad", consta que no fue suprimido sino transitoriamente en 1840, que lo fue de modo definitivo en 1863.

En la página 329, del No 11, se afirma que Sañudo anotó en sus Apuntes, los vicarios de esta Ciudad, "en los siglos XVI y XVII", siendo así que solo apunta los del último. Entre las visitas pastorales, no indica (página 330) las que hubo en el siglo XVIII, en las que sobresale las del obispo de Quito, Nieto Polo de Águila. Finalmente es de saberse que antes del monasterio de la Concepción, los agustinos fundaron el suyo en 1585.

Desde diciembre, el doctor López principió un artículo que llamó "Voces quichuas en Nariño", de modo que creía que pronto propondría una especie de vocabulario de ellas, pero lejos de eso pasaron seis boletines en que solo presentó declinaciones y conjugaciones que se suelen presentar en cualquier gramática del quichua. Hízome recordar pues, la anécdota que trae el Padre Isla que es así: "Había un loco que andaba pregonando por las calles de Sevilla: Cualquier persona que quiera saber cómo se cala un melón, acuda al tío Antón. Llegaban los muchachos y le preguntaban- Tío Antón, corno se cala un melón? ¿Cómo? (respondía el loco en tono muy magistral): Sabiendo el Credo y los Artículos de la Fe". Aquí en el caso del doctor López Alvarez pueden los muchachos preguntar: Tío Antón cuantas son las voces quichuas en Nariño? Y el loco responder- eso se sabe conociendo la declinación del pronombre ñuca.

Para concluir diré, que algunos que se ven impotentes para refutar unas verdades apelan al recurso de sostener que el escrito en que se exponen, es con mala crianza; y que expiación fatalista no puede existir, porque supone libertad. Por eso los brutos animales no expían y Sañudo en la Introducción de sus Estudios, expresamente dijo que forman historia la Providencia divina y la libertad humana, que excluyen todo fatalismo.

Patricio Leal y Franco