

## Análisis de la edición digital de un relato del siglo XIX a través de Voyant Tools

*Analysis of the Digital Edition of a Story from 19th Century through Voyant Tools*

Cristian SUÁREZ-GIRALDO

[suarezg106@gmail.com](mailto:suarezg106@gmail.com)

Universidad EAFIT

### RESUMEN

*Hojas sueltas* fue un tipo de publicación que circuló en Colombia durante la segunda mitad del siglo XVIII e inicios del siglo XIX. Fueron publicaciones cortas y no seriadas que propendían por la divulgación de aspectos diversos de la vida cotidiana, como la cultura, la religión, la economía y la política. “Conversación entre un zapatero y un agrícola de Facatativá sobre la cazería[sic] del Sábado Santo” es un cuento publicado en 1839 en Bogotá por la imprenta José Ayarza, bajo dicho formato de publicación. A través de un análisis de corpus textual, usando la herramienta Voyant Tools, en este artículo se proponen dos mecanismos analíticos: una comparación intertextual e intradiegética de los tres sitios web donde se encuentra alojada la versión digitalizada del cuento y una descripción de líneas discursivas a partir de reseñas críticas elaboradas por estudiantes de literatura. Se concluye que Voyant Tools es una herramienta útil para describir líneas discursivas de un corpus textual, y que puede servir de insumo para realizar análisis del discurso o de tendencias temáticas presentes en las opiniones y reflexiones de un colectivo.

### PALABRAS CLAVE

Análisis textual, Voyant Tools, Edición digital, Narrativa colombiana.

### ABSTRACT

*Hojas sueltas* was a type of publication that circulated in Colombia during the second half of the eighteenth century and the beginning of the nineteenth century. They were short and non-serial publications that promoted the disclosure of various aspects of daily life, such as culture, religion, economy, and politics. “Conversación entre un zapatero y un agrícola de Facatativá sobre la cazería[sic] del Sábado Santo” is a story published in 1839 in Bogotá by the José Ayarza printing press, under that publication format. Through an analysis of the textual corpus, using the Voyant Tools tool, in this article two analytical mechanisms are proposed: an intertextual and intradiegetic comparison of the three websites where the digitized version of the story is hosted and a description of the discursive lines to be found. based on critical reviews prepared by literature students. It is concluded that Voyant Tools is a useful tool to describe discursive lines of a textual corpus and that it can serve as an input to analyze the discourse or thematic trends present in the opinions and reflections of a group.

### KEYWORDS

Textual analysis, Voyant Tools, Digital edition, Colombian narrative.



## 1. INTRODUCCIÓN

*Hojas sueltas* fue un tipo de publicación que circuló en Colombia durante la segunda mitad del siglo XVIII e inicios del siglo XIX. Fueron publicaciones cortas y no seriadas que propendían por la divulgación de aspectos diversos de la vida cotidiana, como la cultura, la religión, la economía y la política. “Dentro de esta denominación se agrupan boletines, cartas, carteles, discursos, separatas de prensa, pasquines, plegables, proclamas, hojas de canciones, ordenanzas, volantes y folletos, entre otros” (Red cultural del Banco de la República [Banrepcultural], s.f., párr. 2).

En diferentes bibliotecas virtuales y repositorios digitales se encuentran versiones digitalizadas (como el Banco de la República, la Biblioteca Miguel de Cervantes, la Biblioteca Luis Ángel Arango, Sala Patrimonial Universidad EAFIT, entre otras) de esta colección de documentos que describen gran parte de las costumbres y modos de pensamiento de la Nueva Granada. En términos generales, dos características sobresalen en el archivo y disponibilidad de este tipo de documentos: la digitalización de los documentos se ofrece en archivos PDF y JPG, lo que dificulta la captura de la información de los contenidos allí registrados; al tratarse de colecciones tan extensas, pocos documentos han sido abordados con detalle o incluidos dentro de un estudio sistemático, lo que hace difícil comprender los contextos de creación y recepción que los enmarcan.

A partir de esta evidencia, en el marco de un curso de Edición digital, del pregrado de Literatura de la Universidad EAFIT, se llevó a cabo entre los meses de octubre y noviembre del 2020 una edición digital del texto “Conversación entre un zapatero y un agrícola de Facatativá sobre la cazería[sic] del Sábado Santo”, un cuento publicado en 1839 en Bogotá por la imprenta José Ayarza, bajo el formato de Hojas sueltas. En el presente trabajo se pretende exponer el resultado del análisis del corpus textual de dicha edición a través del uso de la herramienta Voyant Tools<sup>1</sup>.

## 2. DESCRIPCIÓN DEL CORPUS

El corpus<sup>2</sup> de datos analizado a continuación está compuesto por dos elementos principales: por un lado, el cuento, que fue preparado para la edición digital a través del estándar TEI (*Text Encoding Initiative*), cuyas características estilísticas y formales rozan entre un relato literario y un sermón religioso sobre ciertas actividades que se realizan en medio de la fiesta litúrgica más importante del catolicismo; y, por otro lado, seis reseñas críticas elaboradas por los estudiantes ante la poca información disponible sobre la obra.

<sup>1</sup> Accesible desde: <https://voyant-tools.org/>.

<sup>2</sup> El sitio web donde se aloja este corpus es: <https://suarezg106.gitlab.io/conversacion-eafit/>.

<sup>3</sup> Nos referimos a la Biblioteca Virtual del Banco de la República (<https://babel.banrepcultural.org/digital/collection/p17054coll10/id/682>), la Biblioteca Virtual Miguel de Cervantes (<http://www.cervantesvirtual.com/obra/conversacion-entre-un-zapatero-y-un-agricola-de-facatativa-sobre-la-cazeria-del-sabado-santo-856343/>) y el Repositorio Institucional de la Universidad EAFIT (<https://repository.eafit.edu.co/handle/10784/762>).

<sup>4</sup> Accesible desde: <https://suarezg106.gitlab.io/conversacion-eafit/pages/cuento/>.

<sup>5</sup> Accesible desde: <https://suarezg106.gitlab.io/conversacion-eafit/resenas-criticas>.

### 3. METODOLOGÍA

#### 3.1. Tipo de trabajo sobre los datos

Por la peculiaridad en términos de poca disponibilidad del relato, en esta reseña se llevan a cabo dos tipos de acciones analíticas (integradas a su vez por diferentes variables que se especificarán en el siguiente subapartado):

1. Una comparación intertextual e intradieética: consiste en la revisión sinóptica de los tres sitios web<sup>3</sup> donde se encuentra alojada la versión digitalizada del documento respecto de la edición digital llevada a cabo en el curso de la Universidad EAFIT. Se evalúan las tendencias y el sumario para evidenciar las reiteraciones o ausencias del contenido semántico que cosecha o indexa cada sitio. Además, se revisa la nube de palabras para evidenciar si de los términos más reiterativos pueden extrapolarse algunos motivos literarios que den cuenta de la temática abordada en el cuento o sobre el propósito discursivo que allí se expone.

2. Descripción de líneas discursivas: a partir de las reseñas críticas<sup>5</sup> elaboradas por los estudiantes, se describen cuáles fueron los enfoques de reflexión usados por ellos para completar la edición digital del cuento. Por medio del sumario de documentos se comparan las reseñas para identificar el número total de palabras y el número de palabras únicas en contraste con las nubes de palabras de cada una de las reseñas. Esta contrastación permite señalar posibles líneas teóricas que subyacen a la reflexión hecha por cada estudiante.

#### 3.2. Consideración de variables para el análisis del corpus con Voyant Tools

Antes de exponer los resultados del análisis, a continuación, se presentan algunas variables que se tuvieron en cuenta como diferenciales para el tratamiento de los datos y su comparación.

##### 3.2.1. Primer conjunto de variables

El primer conjunto de variables está relacionado con los diferentes sitios web donde está alojado el relato.

Variable 1.1. En primer lugar, se llevó a cabo el procesamiento de los cuatro sitios donde se encuentra disponible la versión digitalizada del relato a partir de sus URL. La comparación en este nivel es limitada porque Voyant Tools no discrimina los datos extraídos a través de las URL entre componentes textuales propios del relato y los componentes del lenguaje de programación (figura 1). De esta manera el *cirrus* y la tabla de *tendencias* clasificaron términos del diseño de cada página web, en

<sup>6</sup> Accesible desde: <https://voyant-tools.org/?corpus=8cead2027d8873b600fd0ae2cfe3441a&lang=es>.

<sup>7</sup> Accesible desde: <http://www.cervantesvirtual.com/obra/convencion-entre-un-zapatero-y-un-agricola-de-facatativa-sobre-la-caceria-del-sabado-santo-856343/>.

<sup>8</sup> Accesible desde: <https://suarezg106.gitlab.io/convencion-eafit/pages/cuento/>.

tres casos, y en uno, incluso, no se reconoció la página, solicitando *habilitar Javascript* para ello. Términos como *biblioteca*, *virtual*, *pdf*, *view*, entre otros, tienen una frecuencia relativa mayor que términos como *zapatero* y *facatativá*, elementos propios del contenido del relato.

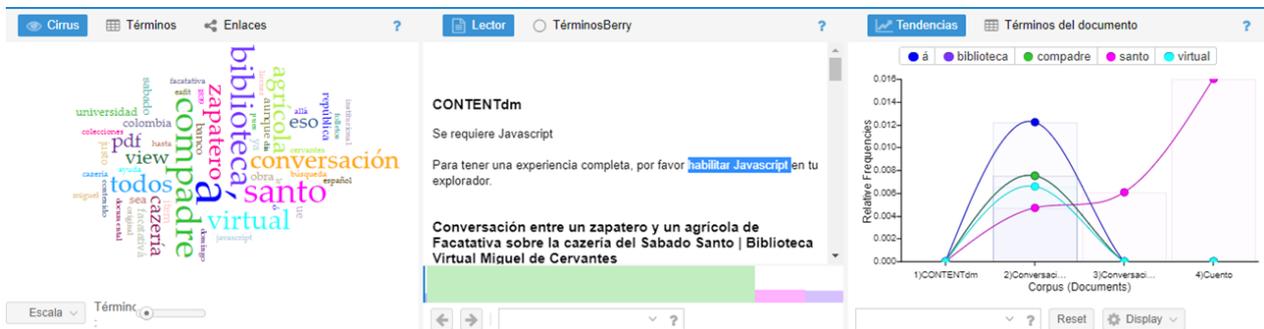


Figura 1. Variable 1.1: resultados y limitaciones. Elaboración propia con Voyant Tools.

Variable 1.2. Después de evidenciar la limitación anterior, se optó por refinar la comparación a los dos sitios web que ofrecen el relato en lenguaje HTML (o lenguaje de marcado para la elaboración de páginas web): el de la Biblioteca Virtual Miguel de Cervantes<sup>7</sup> (Texto A) y el del curso de la Universidad EAFIT<sup>8</sup> (Texto B). En este nivel, Voyant Tools recupera el corpus textual de manera óptima y precisa mejor las características básicas del conjunto de textos. En la variable se evidencia que, aunque el Texto A tiene mayor extensión del documento, densidad del vocabulario y promedio de palabras por oración, el Texto B se compone a su vez de un mayor número de palabras diferenciadas, que corresponden a términos propios del relato (figura 2). Esto se debe no solo a que la herramienta captura la información total de la página del Texto A, incluyendo encabezados y datos de clasificación bibliográfica, sino también a que la edición del Texto B es más cuidada y regularizada respecto del facsimilar que sirvió de base para el etiquetado digital.

- Promedio de palabras por oración:
- Más alto: [Conversación entre un...](#) (44.4)
  - Más bajo: [Conversación entre un...](#) (33.9)
- Palabra más frecuente en el corpus: **no** (49); **á** (18); **compadre** (17); **han** (14); **á** (13)
- Palabras diferenciadas (comparado con el resto del corpus):
1. [Conversación entre un...](#): **á** (13), **e** (9), **biblioteca** (8), **virtual** (7), **í** (6).
  2. [Conversación entre un...](#): **día** (3), **dígame** (2), **días** (2), **cazador** (2), **volvemos** (1).

Figura 2. Variable 1.2: comparación entre Texto A y Texto B. Sección *Sumario*. Elaboración propia con Voyant Tools<sup>9</sup>.

Lo anterior también se puede constatar en el lector, donde podemos comparar las características textuales de cada edición; y en la tabla de tendencias (figura 3), donde se evidencia cómo el Texto A contiene caracteres y palabras que no aparecen en el Texto B, o cómo en el Texto B se

<sup>9</sup> Accesible desde: <https://voyant-tools.org/?corpus=cd8f8b1e8b074bd4268f5e2c0148347e&lang=es>.

<sup>10</sup> Accesible desde: <https://voyant-tools.org/?corpus=cd8f8b1e8b074bd4268f5e2c0148347e&lang=es>.

tomaron algunas decisiones de regularización del lenguaje en su formato HTML. Un caso puntual que ejemplifica lo anterior es la ausencia absoluta de la vocal á (forma vocálica presente en el facsimilar, pero en desuso) en el Texto B, pero frecuente en el Texto A; de igual manera, el uso de la vocal a se incrementa en el Texto B.

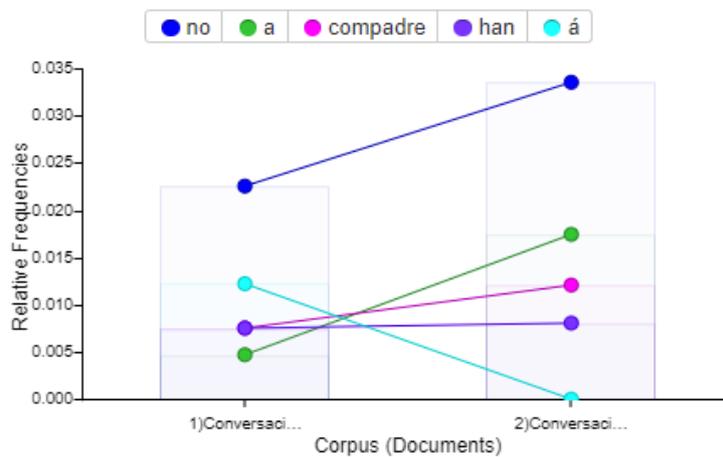


Figura 3. Variable 1.2: comparación entre Texto A y Texto B. Sección *Tendencias*. Elaboración propia con Voyant Tools<sup>10</sup>.

Variable 1.3. Finalmente, se hizo un análisis exclusivo para el sitio donde se encuentra alojado el corpus objeto. El propósito de esta variable es verificar algunos componentes propios de la diégesis del relato a través de las herramientas que ofrece Voyant Tools. Por diégesis entendemos aquí lo que Souriau (1998) define como “el universo de la obra, el mundo establecido por una obra de arte” (p. 445), y lo que a su vez Genette (1989) complementa como “los diversos tipos de representación del discurso de personaje, los modos de presencia explícita del narrador y del lector en el relato” (p. 85). En nuestras palabras, la diégesis es la cualidad proposicional del relato que da sentido a la historia y que, por eso, permite ser seguida como una historia (desde el inicio hasta su desenlace).

Voyant Tools permite un análisis por segmentación, es decir, esta herramienta aísla los términos para medir frecuencias y reiteraciones, y ofrecer algunas ideas del conjunto del discurso que compone el corpus. Respecto al relato analizado, se pudo constatar que los términos más reiterativos que aparecen en el cuadro de tendencias (*a*, *aquí*, *compadre*, *no*, *qué*) (figura 4), así como la nube de palabras, no alcanzan a reflejar la temática del relato. El cuadro de contextos, donde se evidencia la ocurrencia de cada palabra clave con un poco de texto circundante, tampoco expresa esa secuencia proposicional a la que se refiere el componente intradieético.

<sup>11</sup> Accesible desde: <https://voyant-tools.org/?corpus=e2b3b52bb25ca6c0bd29b557e97de02d&lang=es>.

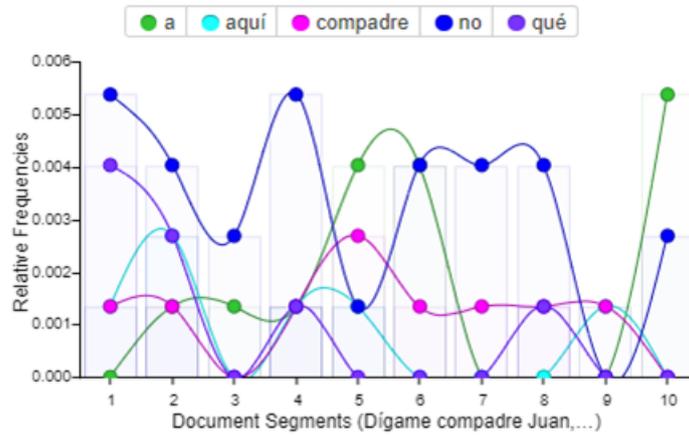


Figura 4. Variable 1.3: análisis exclusivo de la edición realizada por la Universidad EAFIT. Sección *Tendencias*. Elaboración propia con Voyant Tools<sup>11</sup>.

Se podría pensar que este resultado se da por la brevedad del texto, compuesto únicamente por un total de 745 palabras, cuya densidad del vocabulario resultó ser de 0.438; o que Voyant Tools no permite un análisis de obras literarias, que por sus particularidades formales no permitirían conclusiones relevantes frente a una exploración de tipo semántica como la que se realiza a través de esta herramienta.

### 3.2.2. Segundo conjunto de variables

El segundo conjunto de variables está relacionado con las reseñas críticas que elaboraron los estudiantes que participaron de la edición digital del relato.

Variable 2.1. Esta variable integra las seis reseñas críticas procesadas en Voyant Tools después de cargarlas como archivos (TXT) independientes. El conjunto de datos que se representan en el *sumario* nos permite evidenciar que, en promedio, los textos guardan una proporción media respecto a la extensión de los documentos y una densidad del vocabulario que oscila entre 0.582 y 0.406 (figura 5). Este resultado es consecuente con el nivel de estudio en el que se encuentran los estudiantes y el nivel de desarrollo y profundidad que alcanzaron respecto al tiempo en que duró el proyecto de edición digital. Además, es interesante evidenciar que entre las palabras diferenciadas se hallan términos como *lingüística*, *cultura*, *sociedad*, que refieren un tipo de acercamiento peculiar sobre la obra.

<sup>12</sup> Accesible desde: <https://voyant-tools.org/?corpus=977c6699d65ccfd0b1ba834c2522780&lang=es>.

<sup>13</sup> Accesible desde: <https://voyant-tools.org/?corpus=be675549aed6badfaf4fbde118b3f23d&lang=es>.

<sup>14</sup> Accesible desde: <https://voyant-tools.org/?corpus=2f61b9b2cfe00dcd57b9559f1beaa096&lang=es>.

<sup>15</sup> Accesible desde: <https://voyant-tools.org/?corpus=1f45b6e04b087aac32a53cf3fbad85ef&lang=es>.

<sup>16</sup> Accesible desde: <https://voyant-tools.org/?corpus=c7691e1b4fad88bd4af99b7ff4e4d6d9&lang=es>.

<sup>17</sup> Accesible desde: <https://voyant-tools.org/?corpus=b770164c87d94fab20d7569fc603562a&lang=es>.

<sup>18</sup> Accesible desde: <https://voyant-tools.org/?corpus=cc1a7f085ce34d902a470fc7f6173435&lang=es>.



Figura 5. Variable 2.1: cotejo de las reseñas realizadas por estudiantes de la Universidad EAFIT. Sección *Sumario*. Elaboración propia con Voyant Tools<sup>12</sup>.

Variable 2.2. Ahora bien, después de hacer una revisión contrastada de las reseñas, se procedió a una revisión individual de ellas a través de los documentos reseñados como: Reseña 1<sup>13</sup>, Reseña 2<sup>14</sup>, Reseña 3<sup>15</sup>, Reseña 4<sup>16</sup>, Reseña 5<sup>17</sup> y Reseña 6<sup>18</sup>. Sobre cada una se filtraron las palabras vacías (artículos, preposiciones, interjecciones, etc.) y se contrastaron las nubes de palabras (figura 6). En la comparación se evidenció que cada estudiante enfatizó en un elemento particular sobre el texto. Así, las reseñas 1 y 6, al enunciar términos recurrentes como *lingüística*, *habla*, *voz*, *personajes*, *lenguaje*, *texto* y *forma*, acentúan en una reflexión cercana a una concepción formal y compositiva sobre la obra; las reseñas 2, 3 y 4, al usar términos como *sociales*, *vida*, *actividad*, *hombres*, *lector* y *produce*, evidencian un acercamiento más sociológico o hermenéutico; la reseña 5 recurre a usar los mismos términos del cuento, por lo que se puede inferir que lleva a cabo una descripción del relato y no tanto una reflexión propia. En comparación con el currículo del pregrado de Literatura de la Universidad EAFIT, cuyo enfoque principal es hermenéutico y de análisis sobre los géneros literarios, se puede inferir que los estudiantes apelaron a esa formación teórica como recurso para las reseñas críticas.



Genette, G. (1989). *Figuras III*. Lumen.

Red cultural del Banco de la República [Banrepcultural] (s.f.). *Colección de Hojas Sueltas*. <https://www.banrepcultural.org/coleccion-bibliografica/especiales/hojas-sueltas>

Sinclair, S. y Rockwell, G. (2021). *Voyant Tools*. <https://voyant-tools.org/>

Souriau, É. (1998). *Diccionario Akal de estética*. Akal.

Vélez Cardona, A., Euse Lombo, M. C., Mejía Vélez, M. J., Osepa-Mazo, M.-G., Del Valle Uribe, M., Zuluaga Correa, S. y Suárez-Giraldo, C. (Eds.) (2020). *Edición digital de “Conversación entre un zapatero y un agrícola de Facatativá sobre la cazería del Sábado Santo”*. <https://suarezg106.gitlab.io/conversacion-eafit/>