

## Voyant Tools en su aplicación al estudio de las novelas de Rosalía de Castro (1859-1881)<sup>1</sup>

*Voyant Tools in its Application to the Study of Rosalía de Castro's Novels (1859-1881)*

Laura CAMINO PLAZA  
[lcaminoplaza@gmail.com](mailto:lcaminoplaza@gmail.com)

Universidade de Santiago de Compostela

### RESUMEN

Voyant Tools es una herramienta muy útil en el estudio distante de textos literarios, que permite realizar de manera simultánea análisis interpretativos refinados, recuentos cuantitativos y visualizaciones. En este trabajo se mostrarán algunas de estas funcionalidades tomando como campo de pruebas la producción narrativa de Rosalía de Castro (1837-1885), una figura icónica de la cultura gallega y una de las autoras más influyentes de la España decimonónica.

### ABSTRACT

Voyant Tools is a very useful tool for the distant study of literary texts, allowing refined interpretative analyses, quantitative recounts and visualizations to be carried out simultaneously. This paper will show some of these functionalities taking as a testing ground the narrative production of Rosalía de Castro (1837-1885), an iconic figure of Galician culture and one of the most influential authors of nineteenth-century Spain.

### PALABRAS CLAVE

Humanidades Digitales, Voyant Tools, literatura decimonónica, Rosalía de Castro.

### KEYWORDS

Digital Humanities, Voyant Tools, nineteenth-century literature, Rosalía de Castro.



## 1. LA OBRA DE ROSALÍA DE CASTRO DESDE UNA MIRADA RENOVADORA: INTRODUCCIÓN

Este trabajo se enmarca dentro del proyecto de investigación “Cartografías del afecto y usos públicos de la memoria: un análisis geoespacial de la obra de Rosalía de Castro”, que se desarrolla en la Facultad de Filología de la Universidad de Santiago de Compostela y que está dedicado a estudiar la intersección entre memoria, afecto y lugar en la obra rosaliana, concediéndole gran relevancia a la elaboración de cartografías literarias digitales. Desde una perspectiva teórico-crítica, el proyecto pretende desarrollar nuevas propuestas metodológicas y analíticas con las que seguir avanzando en el estudio de la memoria cultural (Rábade Villar, 2018; Labrador Méndez y Rábade Villar, 2019), las relaciones afectivas entre sujetos y lugares (García Candeira, 2018; Camino Plaza, 2020) y las limitaciones y potencialidades de la representación cartográfica en los estudios literarios y culturales (Cabo Aseguinolaza, 2020, 2021), principalmente.

El análisis de corpus que aquí presentamos, elaborado con Voyant Tools, sigue el mismo objetivo básico que el resto de las investigaciones que se encuadran dentro de este proyecto: el de proporcionar una lectura renovada de la obra de Rosalía de Castro, una de las figuras más canónicas del campo literario y cultural gallego.

## 2. LAS NOVELAS DE ROSALÍA DE CASTRO EN VOYANT TOOLS

Para este ensayo hemos utilizado Voyant Tools<sup>2</sup>, una aplicación de código abierto desarrollada por dos investigadores residentes en Canadá: Stéfán Sinclair, de la McGill University, y Geoffrey Rockwell, de la University of Alberta. Con Voyant se pueden llevar a cabo interesantes trabajos de investigación basados en el análisis de corpus. Se usa desde la propia web, sin necesidad de descargar ningún programa, y su uso resulta fácil e intuitivo. Para quien comience desde cero con esta herramienta, es recomendable la consulta de las guías de uso disponibles en el sitio de la aplicación y el tutorial en español que Gutiérrez De la Torre (2019) realizó para *The Programming Historian*. En las siguientes líneas, mostraré algunas de las funcionalidades<sup>3</sup> de Voyant

---

<sup>1</sup> Este trabajo se vincula al proyecto financiado por FEDER / Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación / Proyecto “Cartografías del afecto y usos públicos de la memoria: un análisis geoespacial de la obra de Rosalía de Castro” (FFI201782742-P).

<sup>2</sup> Accesible desde: <https://voyant-tools.org/>.

<sup>3</sup> Las posibilidades analíticas con Voyant son muy variadas y en este trabajo solo presentaremos algunas de ellas. Para consultar todas las herramientas, remitimos a <https://voyant-tools.org/docs/#!/guide/tools>.

para la lectura y el análisis textual tomando como punto de partida el conjunto de novelas escritas por Rosalía de Castro: *La hija del mar* (1859), *Flavio* (1861), *Ruinas* (1866), *El caballero de las botas azules* (1867) y *El primer loco* (1881). Nos centraremos primero en cuestiones generales sobre el corpus y las visualizaciones más básicas que permite Voyant, para después analizar con más detalle las posibilidades que este programa nos ofrece a la hora de estudiar los afectos presentes en las novelas rosalianas.

### 3. REFLEXIONES GENERALES SOBRE EL CORPUS Y VISUALIZACIONES BÁSICAS

Una vez cargadas todas las novelas en Voyant Tools (en texto plano, .txt), en la parte inferior izquierda del área de trabajo podremos ver un resumen general de los datos que el programa ha extraído de nuestro corpus. Tomadas todas las novelas en conjunto, las palabras más frecuentes son: *era* (con 680 reiteraciones), *Flavio* (con 657), *corazón* (515), *Mara* (484) y *entonces* (462). En cuanto a la frecuencia inversa de términos en los documentos (o *tf-idf* por sus siglas en inglés), es decir, las palabras más distintivas de cada novela en comparación con las restantes obras del conjunto, los resultados son los que aparecen recogidos en la parte inferior izquierda de la figura 1. Como es lógico, en la mayoría de los casos las palabras diferenciadoras son los antropónimos de cada obra. Sin embargo, también tenemos algunas pistas sobre la particularidad de *El caballero de las botas azules* (el protagonista no es nombrado sino como duque de la Gloria, vestido siempre con sus particulares botas) o sobre el impulso viajero del protagonista de *Flavio*.



<sup>4</sup> Puede acceder a los datos desde: <https://voyant-tools.org/keywords-d0524c9faa61757f01d7cabdd842b931>.

<sup>5</sup> La densidad se obtiene dividiendo el número de palabras tipo o únicas entre el número total de palabras.

<sup>6</sup> Accesible desde: <https://bit.ly/3EySeEr>.



cuenta todo el corpus novelístico rosaliano, junto con un gráfico de tendencias. Asimismo, es posible personalizar la búsqueda por palabras clave, tal y como se observa en la parte inferior de la figura 4.

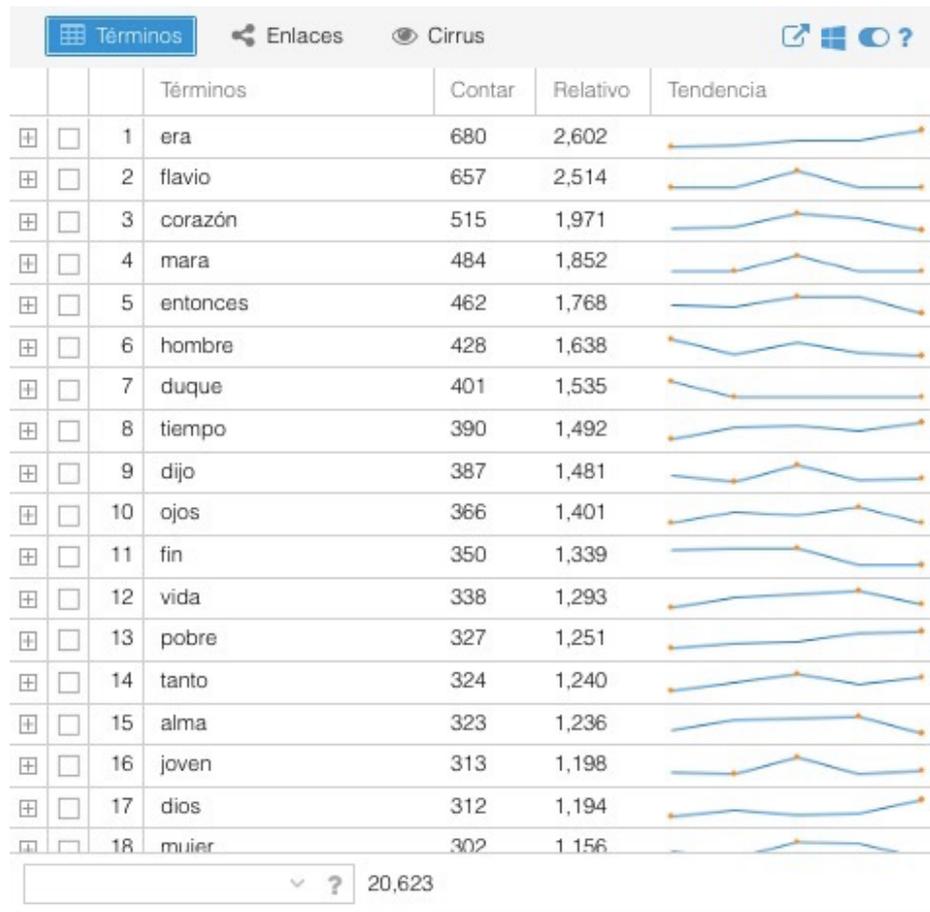


Figura 4. Frecuencias absolutas y relativas de los términos dentro del corpus. Elaboración propia con Voyant Tools<sup>9</sup>.

Si elegimos un término concreto para ver su distribución a lo largo del corpus –en este caso, *duque*– podremos conseguir rápidamente mucha información interesante (figura 5). El sustantivo *duque* aparece exclusivamente en *El caballero de las botas azules* un número total de 401 veces, siendo su frecuencia relativa dentro de esa obra muy importante (5.154 puntos), ya que, como se vio anteriormente, es el término que la autora eligió para definir al misterioso protagonista. Pero si aplicamos una búsqueda sintáctica, veremos que el lexema *duque*\* nos arroja un resultado de 408 apariciones totales, puesto que a lo largo del corpus rosaliano aparecen también los términos flexionados *duques* (4 veces en *El caballero de las botas azules*, con una frecuencia relativa baja, de 51 puntos), *duquesa* (1 vez en *El caballero de las botas azules*, con una frecuencia relativa de 13 puntos, y otra vez en *Ruinas*, con una frecuencia relativa de 48), así como *duquesas* (1 vez en *El caballero de las botas azules*, con una frecuencia relativa de 13 puntos).

<sup>9</sup> A este respecto, véase el análisis ofrecido por R. Carvalho Calero en su trabajo “Falta um capítulo em *El caballero de las botas azules*”, publicado en 2001 en *Nova Renascença* y reproducido por P. García Negro en 2019 para la revista *Follas Novas*.



Figura 5. Frecuencias absolutas y relativas de *duque\**, *duque*, *duques*, *duquesa* y *duquesas* en el corpus. Elaboración propia con Voyant Tools

Efectivamente, si utilizamos la herramienta *Contextos* (figura 6), podremos comprobar la presencia de los términos ahora desglosados a lo largo del corpus, pudiendo acceder a los pasajes donde aparecen mencionados. Así, por ejemplo, observaremos que *duquesa* está siendo utilizado en el documento 1 del corpus de trabajo, es decir, *El caballero de las botas azules*, para referirse a la duquesa de Río Ancho. Este hecho guarda correspondencia con el afán de Rosalía de Castro por describir en esta novela una gran variedad de tipos femeninos, especialmente de clase alta, aunque en ocasiones sea a través de menciones fugaces, como ocurre en este caso<sup>9</sup>. La otra referencia a una duquesa, totalmente despersonalizada, se encuentra en el documento 5, es decir, en *Ruinas*.

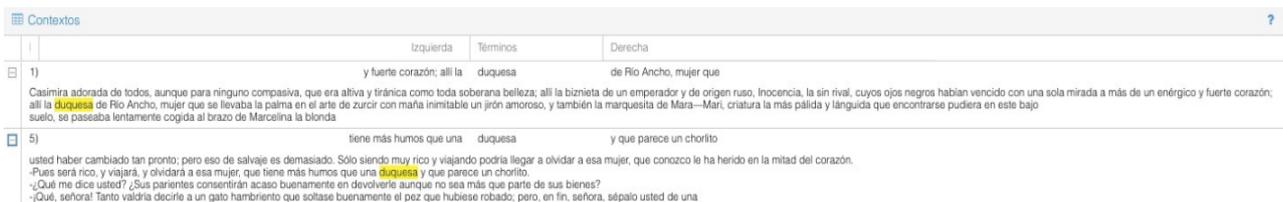


Figura 6. El término *duquesa* en el corpus rosaliano. Elaboración propia con Voyant Tools.

Otra herramienta con la que Voyant permite explorar las frecuencias relativas entre términos y documentos es *Mandala* (figura 7).

<sup>10</sup> Accesible desde: <https://voyant-tools.org/?corpus=161a17a6821a66a58be1ffde60495112&lang=es&stoplist=keywords-d0524c9faa61757f01d7aabdd842b931&query=era&query=flavio&query=coraz%C3%B3n&query=maro&query=entres&query=harbre&query=dque&query=tempo&query=cp&query=cp&view=Mandala>





Figura 9. Co-presencia del término *Mara* en *El caballero de las botas azules* y *Flavio*. Elaboración propia con Voyant Tools.

#### 4. ANÁLISIS AFECTIVO CON VOYANT TOOLS

Desde el punto de vista de los afectos, uno de los nodos conceptuales del proyecto de investigación en el que se encuadra este trabajo, Voyant también resulta una primera herramienta de análisis *distante*<sup>11</sup> provechosa.

En la figura 10, vemos cómo la búsqueda de la frecuencia de la familia léxica *trist\** y *alegr\** (la cual nos permite capturar la presencia de términos como *triste*, *tristeza*, *tristemente*, *entristecen...* o de *alegría*, *alegre*, *alegremente*, etc. a lo largo de las obras) nos arroja el siguiente gráfico de tendencias (figura 11), donde se constata que la novela con mayor presencia de vocablos léxicamente relacionados con la emoción *tristeza*<sup>12</sup> es *La hija del mar*.

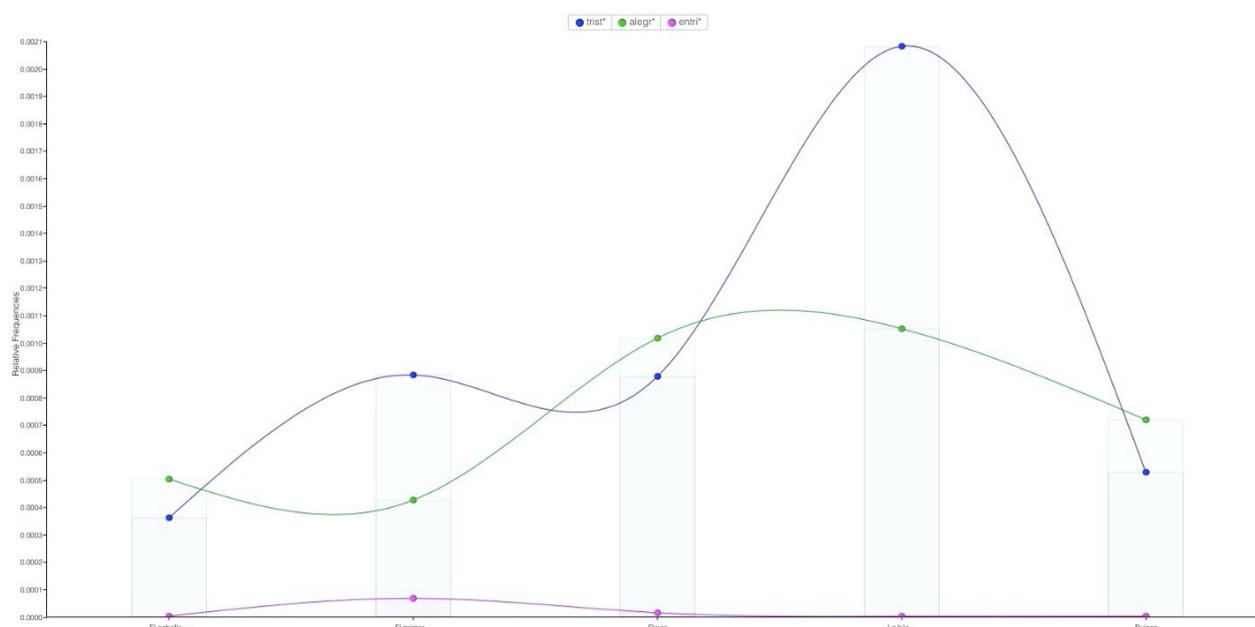


Figura 10. Familia léxica de *trist\** y *alegr\**. Elaboración propia con Voyant Tools.

Figura 11. Estudio comparado preliminar de la presencia de los afectos *tristeza* y *alegría* en las novelas rosalianas. Elaboración propia con Voyant Tools<sup>13</sup>.

Ahora que sabemos que *La hija del mar* es la obra con mayor presencia de la familia léxica relacionada con la emoción *tristeza*, resulta útil investigar cómo se distribuyen las menciones a este afecto a lo largo de la narración.

La herramienta *Bubblelines* permite capturar la distribución de un término o más a lo largo de un documento o de un conjunto de textos, y posibilita su visualización ordenada (figura 12). Cada texto es dividido por el programa en fragmentos de igual longitud y la visualización resultante indica, mediante diferentes tamaños de burbujas, cuál es la frecuencia de una palabra en un seg-

14 Accesible desde: <https://voyant-tools.org/bdd809e32db1abdea09ba11980b81760&view=Bubblelines>.

mento determinado del documento. Pasando el cursor por encima de las burbujas, el usuario puede contabilizar el número de apariciones de cada término:



Figura 12. Distribución ordenada de los afectos *tristeza* y *alegría* en cada novela. Elaboración propia con Voyant Tools<sup>14</sup>.

Como se puede inferir del resultado obtenido, la presencia de vocablos relacionados con la tristeza en *La hija del mar* se condensa desde la mitad hacia el final de la novela. Tras una complicada existencia, marcada por el abuso de su propio padre y la imposibilidad de disfrutar de su amor verdadero, Esperanza, hija adoptiva de Teresa, acabará suicidándose en el mar, del que había sido rescatada cuando era una niña. Desde el capítulo XIV, titulado “El entierro”, en el que se da sepultura a Fausto, el amado de Esperanza, la presencia de la tristeza se dispara. Si en esta parte de la novela el lexema *trist\** aparece 9 veces, en el siguiente capítulo, llamado “La loca”, en el que se relata la locura y profundo pesar en los que se sumió Esperanza, se contabilizan 10 apariciones, seguidas de otras 5 en el capítulo XVII (“Zozobra”), 3 en el capítulo XVIII (“Crisis”) y otras 3 en el capítulo XIX (“La razón”). El clímax llega en el último capítulo, titulado “Justicia de Dios”, donde Alberto Ansot, padre de Esperanza, es ajusticiado por sus crímenes. En esta parte el lexema *trist\** aparece 23 veces. Por último, en la conclusión de la novela, que acaba con el suicidio de Esperanza y el lamento de Teresa, se contabilizan otras 3 apariciones. A la luz de estos datos, se puede concluir que el arco emocional de *La hija del mar*, si seguimos la clasificación dada por Reagan *et al.* (2016), se encuadraría dentro del esquema llamado “Ícaro”, donde el pico positivo del inicio de la novela se contrapone con su dramática caída final.

Puede ser interesante compaginar esta radiografía general del léxico de *La hija del mar* con una aproximación más cercana. Gracias a la herramienta *Corpus Collocate* (figura 13) podremos estudiar en su contexto los términos clave que hayamos previamente seleccionado:

| Colocaciones             |         |                      |                   |
|--------------------------|---------|----------------------|-------------------|
|                          | Término | Contexto del término | Contar (contexto) |
| <input type="checkbox"/> | trist*  | teresa               | 6                 |
| <input type="checkbox"/> | trist*  | mirada               | 6                 |
| <input type="checkbox"/> | trist*  | era                  | 6                 |
| <input type="checkbox"/> | trist*  | soledad              | 5                 |
| <input type="checkbox"/> | trist*  | silencio             | 5                 |
| <input type="checkbox"/> | trist*  | pobre                | 5                 |
| <input type="checkbox"/> | trist*  | algunas              | 5                 |
| <input type="checkbox"/> | alegr*  | rayo                 | 5                 |
| <input type="checkbox"/> | trist*  | preguntó             | 4                 |
| <input type="checkbox"/> | trist*  | palabras             | 4                 |
| <input type="checkbox"/> | trist*  | añadió               | 4                 |
| <input type="checkbox"/> | trist*  | alberto              | 4                 |
| <input type="checkbox"/> | trist*  | aire                 | 4                 |
| <input type="checkbox"/> | trist*  | acento               | 4                 |

Figura 13. Colocaciones más frecuentes de los lexemas *trist\** y *alegr\** en *La hija del mar*. Elaboración propia con Voyant Tools.

## 5. CONCLUSIONES Y FUTURAS VÍAS DE TRABAJO CON VOYANT TOOLS

Voyant Tools es una herramienta de investigación muy útil, que resulta de gran provecho a la hora de combinar la lectura distante de un documento o corpus con su lectura atenta, tradicional. En muchas ocasiones, la combinación de un estudio cuantitativo de los datos con un análisis cualitativo nos permitirá constatar hipótesis de investigación previas; pero, en muchas otras, Voyant podrá arrojarnos nuevas conclusiones o guiarnos hacia hipótesis no contempladas anteriormente.

A la luz de mi experiencia como usuaria de Voyant Tools en su aplicación al corpus novelístico de Rosalía de Castro, considero que una interesante vía para futuras exploraciones puede ser la división de cada obra por capítulos, sobre los que luego ejecutar análisis individuales pormenorizados. Esto me permitiría constatar la presencia de ciertos términos afectivos a lo largo de cada narración de una manera más detallada y ordenada, lo que posibilitaría la identificación de los picos emocionales de cada novela. Se abre, de este modo, una prometedora vía de análisis de corpus, que complementará beneficiosamente los estudios de corte tradicional que siguen proliferando alrededor de la obra y figura de Rosalía de Castro.

## REFERENCIAS BIBLIOGRÁFICAS

- Cabo Aseguinolaza, F. (2020). Lugar y mapas: ensayo cartográfico sobre “Orillas del Sar”, de Rosalía de Castro. *Tropelías. Revista de Teoría de la Literatura y Literatura Comparada*, (7), 842-853.
- Cabo Aseguinolaza, F. (2021). De espacio, literatura y mundo. *Theory Now. Journal of Literature, Critique, and Thought*, 4(1), 22-41. <http://dx.doi.org/10.30827/TNJ.v4i1.16789>
- Camino Plaza, L. (2020). Memoria, género y comunidad en los prólogos de Rosalía de Castro. *Itinerarios. Revista de estudios lingüísticos, literarios, históricos y antropológicos*, (32), 115-129.

<http://itinerarios.uw.edu.pl/memoria-genero-y-comunidad-en-los-prologos-de-rosalia-de-castro/>

De Castro, R. (1859). *La hija del mar*. Imprenta de J. Compañel. [http://www.cervantesvirtual.com/portales/rosalia\\_de\\_castro/obra/la-hija-del-mar--1/](http://www.cervantesvirtual.com/portales/rosalia_de_castro/obra/la-hija-del-mar--1/)

De Castro, R. (1861). *Flavio*. En V. García Martí (Ed.) (1994), *Obras completas, T. II* (pp. 203-473). Aguilar. [http://www.cervantesvirtual.com/portales/rosalia\\_de\\_castro/obra/flavio--0/](http://www.cervantesvirtual.com/portales/rosalia_de_castro/obra/flavio--0/)

De Castro, R. (1866). *Ruinas*. El Museo Universal. [http://www.cervantesvirtual.com/portales/rosalia\\_de\\_castro/obra/ruinas--1/](http://www.cervantesvirtual.com/portales/rosalia_de_castro/obra/ruinas--1/)

De Castro, R. (1867). *El caballero de las botas azules*. Imprenta de Soto Freire. [http://www.cervantesvirtual.com/portales/rosalia\\_de\\_castro/obra/el-caballero-de-las-botas-azules--0/](http://www.cervantesvirtual.com/portales/rosalia_de_castro/obra/el-caballero-de-las-botas-azules--0/)

De Castro, R. (1881). *El primer loco*. Imprenta y librería de Moya y Plaza. [http://www.cervantesvirtual.com/portales/rosalia\\_de\\_castro/obra/el-primero-loco--0/](http://www.cervantesvirtual.com/portales/rosalia_de_castro/obra/el-primero-loco--0/)

García Candeira, M. (2018). A cidade e os seus (des)contentos. Santiago de Compostela na obra de Rosalía de Castro e Emilia Pardo Bazán. *Grial: revista galega de cultura*, 56(220), 40-49.

García Negro, P. (2019). Falta um capítulo em *El Caballero de las botas azules*. Ricardo Carvalho Calero. *Follas Novas: revista de estudos rosalianos*, (4), 160-168.

Gutiérrez De la Torre, S. (2019). Análisis de corpus con Voyant Tools. *The Programming Historian en español*, (3). <https://doi.org/10.46430/phes0043>

Labrador Méndez, G. y Rábade Villar, M. do C. (2019). A memoria alienada. Revolución, biopoder e ecoloxías comunais en *El primer loco* y “El Domingo de Ramos” (1881). En G. Labrador Méndez y M. do C. Rábade Villar (Eds.), *Rosalía de Castro. El primer loco* (pp. 9-82). Alvarellos Editora.

Moretti, F. (2015). *Lectura distante*. Fondo de Cultura Económica.

Rábade Villar, M. do C. (2018). ¿Un asunto de estado? Usos públicos de la memoria literaria. *Tropelías. Revista de Teoría de la Literatura y Literatura Comparada*, (4), 161-185.

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M. y Dodds, P. Sh. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(31), 1-28. <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0093-1>

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