

Dos concordancias virtuales de obras de Quevedo: el *Buscón* y la *Poesía completa*

Alfonso Rey
María José Alonso Veloso
Grupo Quevedo (GI-1373)



II Xeira CLARIAH-GAL
Santiago de Compostela
9 de mayo de 2025

El presente póster describe dos concordancias virtuales de la obra de Francisco de Quevedo: la narración picaresca titulada *Historia de la vida del Buscón llamado don Pablos, ejemplo de vagamundos y espejo de tacaños*, según el texto de la versión B, que parece la más tardía de las cuatro que se conocen (S, C y Z son las otras tres versiones); y la *Poesía completa*, un total de 916 poemas, 891 de autoría segura, y el resto de atribución posible, pero más dudosas.

En el primer caso, se ha hecho un marcado por capítulos y otro por palabras, junto con una generación por colocaciones y frecuencias. La sección alfabética ordena las palabras por su letra inicial, indicando su número de ocurrencias. Al marcar una de ellas se ofrece el contexto en que aparece, así como el capítulo del *Buscón* al que corresponde. La sección Frecuencia ofrece los mismos datos desde otra perspectiva, ordenando las palabras

en función de su frecuencia, según un orden creciente. La frecuencia 1, el Hapax Legomena, incluye las concordancias de todas las palabras que tienen una sola aparición en el texto.

En el caso de la poesía, en el que se ha aplicado una metodología semejante, el marcado ha sido por *musas* (secciones) y por palabras. Además de ofrecerse el contexto en que aparece, se indican el apartado (alguna de las nueve musas, la *musa décima*, el *Heráclito* o los poemas atribuidos), y también el número de poema y verso. Por otra parte, el buscador del Índice de primeros versos permite localizar palabras situadas en los mismos.

Las concordancias se basan, con leves modificaciones, en el texto fijado por la versión publicada por Alfonso Rey en su edición crítica de las cuatro versiones del *Buscón* (Madrid, Consejo Superior de Ediciones Científicas,

págs. 345-439); y en el de las ediciones publicadas por Alfonso Rey y María José Alonso Veloso (*Poesía completa*, Barcelona, Castalia, 2021; *Silvas*, Salamanca, Universidad, 2024).

La literatura y la lengua poética de Quevedo, a quien el español debe tantas innovaciones y hallazgos, constituyen un recurso cultural que debe ser preservado y difundido no sólo en soporte escrito, sino también con ayuda de los desarrollos que aportan las tecnologías del conocimiento en lo que se refiere al tratamiento digital de los textos literarios. Concordancias virtuales como las que aquí se presentan, fruto de casi tres décadas de estudio de la prosa y la poesía quevedescas, hacen visible para el conjunto de la ciudadanía y la comunidad científica un importante caudal de datos filológicamente seguros en los que podrán basarse posteriores investigaciones.

Para el desarrollo de la web quevedo.usc.gal y la visualización de las concordancias de una parte de la obra prosística (el *Buscón*) y la *Poesía completa* de Francisco de Quevedo se realizó:

1. Tratamiento del corpus para extraer la información necesaria en las concordancias.
2. Diseño y desarrollo de la web estática, con información almacenada en archivos html.



1. Tratamiento del corpus y creación de *htmls*

a. Etiquetado y división del documento

En el caso del *Buscón* se partió de un documento en formato txt: fueron etiquetados los títulos de los tres libros y de cada uno de los capítulos, para poder dividir el corpus después mediante la librería *Quanteda* en el lenguaje de programación R. De esta forma se recogió un índice, con cada libro y sus capítulos, y se exportó cada capítulo en una página.

Para las concordancias del conjunto de poemas de *Quevedo*, debido a la disparidad de formas y el volumen de los datos, la división y el etiquetado inicial fueron más complejos. En primer lugar, el documento se dividió y fue etiquetado conforme a las nueve *Musas* de las ediciones póstumas, a las que se añadieron tres apartados finales: la *Musa décima*, *Heráclito cristiano* y *Poemas atribuidos*.

Debido a la presencia de otras divisiones internas, fue necesario crear un código de colores que permitiese resaltar el texto con el color correspondiente, en base a la tipología de textos de cada poema (comentarios, títulos de poema, subtítulos, nombres de personaje, etc.). Se hizo un resaltado previo automatizado mediante macros (scripts ejecutados dentro de *Word*) en el documento. Con el nuevo archivo de texto, se procedió al etiquetado mediante una serie de scripts de *Python* y a la creación de:

- El **texto limpio** con el contenido del poema del que salen las concordancias.
- El **texto como se debe mostrar** en la página de cada poema.
- El **índice** (en formato json).



b. Creación de concordancias, frecuencias y listado alfabético

El proceso con el *Buscón* y la *Poesía completa* fue semejante. En el primer caso se realizó en R con la librería *Quanteda*, y en el caso de los poemas se hizo en *Python* con la librería *nltk*.

El texto limpio del poemario se cargó en el script para que se pudieran hallar las concordancias en todo el texto. El texto de los poemas (palabras y signos de puntuación), se recorrió palabra a palabra y por cada una se hallaron todas sus concordancias en el texto, con el número de palabras anteriores y posteriores establecido previamente. En el caso del *Buscón* fueron las cinco palabras anteriores y posteriores; en los poemas fue necesario restringir el contexto al verso donde se encuentra la palabra. Además, se recogió la información de contexto global: libro y capítulo, para la obra en prosa, y *musa*, poema y número de verso en el caso de la poesía.



Con este método, se creó para cada palabra contenida en la obra (sin repeticiones) un documento *html* con todas sus concordancias numeradas.

En cuanto a las frecuencias, se establecieron los rangos de frecuencia correspondientes, en base al total de palabras y el número de apariciones: En el *Buscón*: 15-20, 21-50, 51-100, 101-200, 201-500, 501-1000, 1001-2000, 2001-2308. En la *poesía completa*: 2, 3, 4, 5, 6, 7, 8, 9, 10 [...] 21-50, 51-75, 76-150, 151-225, [...] 501-1000, 1001- 9126.

Se recorrió todo el corpus palabra a palabra y para cada una se analizó su frecuencia en el corpus y se almacenó. Luego se recorrieron todos los rangos y se añadieron en cada uno todas las palabras que se correspondían ordenadas alfabéticamente. Para cada frecuencia se creó una tabla y se exportó en un documento *html*.

En el caso del listado alfabético del total de palabras se realizó el mismo proceso: para cada letra del abecedario se añadieron las palabras que correspondían con su frecuencia y se exportó lo obtenido en un documento *html*.



2. Desarrollo web

a. Creación de documentos *htmls* del contenido

Para poder mostrar la información correctamente, una vez dispusimos de los *htmls* con los datos estadísticos de concordancias, frecuencias y orden alfabético, se crearon los documentos *html* del contenido propiamente (los poemas o los capítulos).

Así, los documentos etiquetados y en *txt* de cada capítulo o poema se pasaron a un *html* mediante un script en *Python* utilizando la librería *lxml*.

Estos documentos siguieron el diseño pensado para asignar a cada elemento etiquetado la clase correspondiente (cursivas, títulos, etc.). Fue creado un *css* con todos los elementos existentes en los poemas y la tipología que le íbamos a dar.

b. Inclusión de vínculos a los *html*

A continuación, cada *html* de un poema o capítulo se editó para añadir en cada palabra del contenido del poema el enlace correspondiente al documento con las concordancias de la misma.

Estos enlaces se incluyeron asimismo en los *htmls* de cada frecuencia y de cada letra del alfabeto.



c. Pruebas y validación

Antes de añadir todo el contenido a la web se hicieron pruebas mediante *scripts* para confirmar que el número de palabras y enlaces que existen es correcto, que el contenido se corresponde con el que debe ser, que no hay sangrías inadecuadas ni caracteres extraños, etc.



d. Desarrollo en *css* y *javascript* del diseño y funcionamiento de la web

Una vez definido el contenido, se desarrolló también el resto de las páginas que se quieren mostrar, creando documentos *html* y su diseño en *css*: el inicio, los índices (a partir de los archivos *json*), la cabecera y pie, los apartados de "Proyecto", "Obras", "Agradecimientos" y "Como citar", al tiempo que se añadieron los logos y otros textos. Por último, mediante *JavaScript* se creó toda la navegación entre páginas, cambios de capítulo o poema mediante flechas, buscador para los índices de la obra, despletables, etc.



Las concordancias de este proyecto han sido realizadas por Imaxin, Factoría de Software e Multimedia S.L.

Este trabajo ha contado con la financiación de:

Programa de Consolidación y Estructuración de la Xunta de Galicia, referencia ED 431B 2021/005, para el período 2021-2023; y ED431B2024/15 para 2024-2026. Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación. Convocatoria de Proyectos de Generación de Conocimiento (investigación no orientada). «Edición crítica y anotada de la poesía completa de Quevedo, 2. *Las tres musas*», referencia PID2021-123440NB-I00, para el período 2022-2025.



Referencias

- Rey, Alfonso (ed.): *El Buscón: edición crítica de las cuatro versiones*, Madrid, Consejo Superior de Ediciones Científicas, 2007.
- Rey, Alfonso y Alonso Veloso, María José (eds.): *Francisco de Quevedo, Poesía completa*, Barcelona, Castalia Maior, 2021.
- Rey, Alfonso y Alonso Veloso, María José (eds.): *Francisco de Quevedo, Silvas* (edición crítica y anotada), Salamanca, Ediciones Universidad de Salamanca, (colección Textos recuperados), 2024.