

PORTAL DE CORPUS HISTÓRICOS IBERORROMÁNICOS (CORHIBER)

Joan Torruella (*ICREA – UAB*)
Joan.torruella@uab.cat
Código Orcid: 0000-0002-0233-7120

Johannes Kabatek (*Universität Zürich*)
kabatek@rom.uzh.ch
Código Orcid: 0000-0001-8743-6250

1. PRESENTACIÓN

En el tercer congreso dedicado a *Corpus Diacrónicos en Lenguas Iberorrománicas (CoDiLI)*, que tuvo lugar en Zurich en el año 2014, y que había sido precedido por los de Palma de Mallorca (2007) y Barcelona (2011)¹, nos dimos cuenta de que en la Red ya existía, acabados o en proceso de construcción, un número considerable de corpus de carácter histórico relacionados con las lenguas de la Península Ibérica que podían ser muy útiles para el estudio de las lenguas iberorrománicas pero que eran desconocidos por buena parte de los estudiosos del tema. Por ello se pensó en la posibilidad de crear un Portal en Internet donde se encontraran las descripciones básicas de estos corpus y el enlace a su URL (*Uniform Resource Locator*) o dirección electrónica.

En consecuencia, se decidió crear un espacio en la Red (un Portal) denominado CORHIBER (Corpus Históricos IBERorrománicos), el cual es de acceso libre en Internet (<http://corhiber.org>).

2. JUSTIFICACIÓN DEL TÍTULO

El recurso que se ha creado lleva por título «Portal de *Corpus Históricos Iberorrománicos*». Esta denominación se compone de cuatro conceptos que pensamos definen exactamente el contenido de la aplicación.

A) Portal

Según la acepción de *portal* marcada como término informático (*Inform.*) en el *Diccionario de la Lengua Española* de la RAE, «portal» es el «espacio de una red informática que ofrece, de forma sencilla e integrada, acceso a recursos y servicios». Y es

¹ Actualmente se han celebrado cinco congresos *CoDiLI*: Mallorca 2007, Barcelona 2011, Zúrich 2014, Alcalá de Henares 2016 y Gent-Bruselas 2019. El *VI Congreso Internacional de Corpus Diacrónicos en Lenguas Iberorrománicas (CODILI6)*, será los días 5-7 de octubre de 2022 en la *Università Ca' Foscari*, de Venecia.

justamente eso lo que queremos que sea nuestro recurso: un punto en Internet des del que, de forma fácil y amigable, se pueda acceder a los distintos servicios que ofrece cada uno de los recursos (los corpus, en este caso) a los que el Portal apunta.

B) Corpus

El término *corpus* expresa el concepto más matizable de los cuatro que definen nuestro recurso, puesto que, muchas veces, se utiliza de manera poco precisa o, incluso, errónea. Si partimos de una de las definiciones más aceptadas en el mundo de la Lingüística de Corpus que es la que en el año 1991 redactó John Sinclair y que, posteriormente, en el año 1996, hizo suya el grupo EAGLES², un corpus es «una colección de textos, **seleccionados y ordenados de acuerdo con criterios lingüísticos explícitos**³, para ser utilizados como una muestra de la lengua». Sin embargo, dentro del mundo de las colecciones textuales existen al menos tres tipos de recopilaciones según el grado y exigencia en la selección y ordenación de los textos, de los cuales solamente uno responde al concepto de 'corpus' en sentido preciso:

- Archivo digital.- Recopilaciones de textos muy abiertas sin criterios específicos de selección y ordenación.
- Biblioteca electrónica.- Recopilación de textos unificados en un formato estándar, siguiendo ciertas normas de contenido, pero sin un criterio riguroso de selección ni de ordenación.
- Corpus textual.- Recopilación de textos unificados en un formato estándar y seleccionados siguiendo unos criterios de equilibrio y representatividad según una finalidad específica.

En nuestro caso usamos el término «corpus» en su significado más amplio. Así, en el Portal, además de incluir recopilaciones que responden al concepto de 'corpus' en el sentido más preciso del término, también se incluyen recopilaciones textuales menos restrictivas en el cumplimiento de estrictos criterios de selección, como pueden ser archivos digitales, bibliotecas electrónicas, diccionarios, tesoros y repertorios. De esta manera, en *CORHIBER* se presentan:

- Corpus:
 - *CIPM (Corpus Informatizado de Português Medieval)*
 - *CORHIJE (Corpus Histórico Judeoespañol)*
 - *CICA (Corpus Informatizat del Català Antic)*
- Archivos digitales:
 - *ADMYTE (Archivo Digital de Manuscritos y Textos Españoles)*
 - *OSTA (Old Spanish Textual Archive)*
 - *P. S. (Post Scriptum. A Digital Archive of Ordinary Writing)*

² EAGLES, *Expert Advisory Group on Language Engineering Standards*.

³ La negrita es nuestra.

- Bibliotecas electrónicas:
 - *BIDTEA (Biblioteca Digital de Textos del Español Antiguo)*
- Diccionarios:
 - *DiCCA XV (Diccionari del castellà del segle XV a la Corona d'Aragó)*
 - *Diccionario de Textos Concejiles de Andalucía*
 - *Diccionari de Textos Catalans Antics*
- Tesoros:
 - *TILG (Tesouro Informatizado da Lingua Galega)*
 - *TLPGP (Tesouro do Léxico Patrimonial Galego e Português)*
- Repertorios:
 - *RIALC (Repertorio Informatizzato della'Antica Letteratura Catalana)*

Sin embargo, todos estos recursos tienen en común que presentan recopilaciones de textos seleccionados y ordenados con criterios lingüísticos y dan acceso a ellos a partir de un motor de búsquedas léxicas, más o menos sofisticado, que permiten extraer muestras concretas y reales de la lengua.

Además, en la publicación de recursos en el Portal se han hecho algunas excepciones, incluyendo algunos que no se pueden considerar corpus pero que se han estimado útiles para el estudio de las lenguas iberorrománicas, ya sea porque dan acceso a un texto importante ya sea porque dan acceso a bases de datos con informaciones complementarias a los corpus que se recogen en *CORHIBER*. Así, actualmente en el Portal también se da entrada a:

- *La Fazienda de ultramar*, puesto que, a pesar de que se trata de un solo texto — y, por lo tanto, no configura un corpus—, consideramos que, por su importancia dentro del romance pre-alfonsí, puede ser un buen recurso de obtención de datos textuales para los estudiosos de la lengua de esta época.
- *PhiloBiblon*, ya que se trata de una base de datos bibliográfica que por el enorme caudal de datos que contiene sobre textos romances escritos en la península ibérica en la Edad Media y temprano Renacimiento, nos ha parecido útil que su acceso figurara en el Portal.

C) Históricos

Según el diccionario académico «histórico» se refiere a algo «Perteneiente o relativo a la historia», y «historia» se presenta como «1. f. Narración y exposición de los acontecimientos pasados y dignos de memoria, sean públicos o privados. 2. f. Disciplina que estudia y narra cronológicamente los acontecimientos pasados». Es decir, algo para ser histórico debe referirse a acontecimientos del pasado. La cuestión está en estipular hasta qué momento es «pasado» y, también, a partir de qué momento es «pasado».

Es obvio que el año 2020 a día de hoy ya es pasado, que ayer ya es pasado y que, incluso, hace un minuto ahora ya es pasado, es decir, en realidad, mirando hacia atrás, todo es pasado. Ello obliga que nuestro proyecto tenga que definir en qué momento

quiere situar el corte cronológico más próximo de pasado. Sabiendo que se trata de una fecha convencional y que puede ser cambiante, en el proyecto se ha considerado «pasado» hasta el final del siglo XIX. De este modo, cualquier corpus que contenga textos anteriores al siglo XX es candidato a ser incluido en *CORHIBER*. En el futuro, se podría extender el límite temporal más próximo hasta el final del siglo XX, lo que permitiría incluir los corpus orales existentes en el Portal, pero esto, de momento, es solo una posibilidad que se deberá contemplar más adelante.

Más complejo es estipular a partir de cuándo se puede considerar «pasado» de una lengua o, más concretamente, a partir de cuándo se puede asignar un texto a una lengua. Tradicionalmente, la atribución de un texto a una lengua concreta se hace teniendo como base el componente léxico, pero se podrían considerar otros aspectos como, por ejemplo, el sintáctico. Los archivos y bibliotecas están llenos de documentos redactados principalmente con palabras latinas pero con una sintaxis totalmente romance.

Por ello, *CORHIBER* se propone dar cuenta también de aquellos corpus de textos latinos medievales que de algún modo puedan aportar datos interesantes para el estudio de los orígenes de las lenguas iberorrománicas, puesto que, tal como se expresa en la presentación del *Corpus Documentale Latinum Cataloniae*, se trata de «textos que se erigen en privilegiado testimonio escrito de la lengua latina medieval, pero también de la incipiente lengua románica».

D) Iberorrománicos

En *CORHIBER* partimos del concepto iberorrománico que se expresa en la segunda acepción del diccionario académico, «Grupo de lenguas romances derivadas del latín hablado en el antiguo territorio de Iberia o Hispania, entre las que destacan el español, el portugués, el gallego y el catalán». Es decir, en el Portal se tienen en cuenta todas las lenguas que se han hablado y escrito en la Península Ibérica derivadas del latín, y las organizamos según las cuatro que la propia definición remarca como destacadas: español, portugués, gallego y catalán. Sin embargo, además, se han creado dos nuevos apartados, uno dedicado al judeoespañol y otro a los corpus latino-romances.

3. ESTRUCTURA DEL *CORHIBER*

El *Portal CORHIBER* se estructura a partir de dos ejes, uno de principal (lenguas) y otro secundario (ámbitos).

Por lo que se refiere al eje principal, el de lenguas, como ya se ha indicado, éste se estructura a partir de las cuatro grandes lenguas iberorrománicas de la Península Ibérica: portugués, gallego, español y catalán. Además, se ha incluido un apartado dedicado al judeoespañol y otro al latino-romance. También en el mismo eje de lenguas se ha incluido un apartado dedicado a los corpus plurilingües, es decir, a los corpus que contienen textos de distintas lenguas. Así, tenemos el *Dialogyca BDDH* dedicado a las lenguas peninsulares, *Post Scriptum* dedicado a cartas redactadas en español i en portugués y el *Tesouro do léxico patrimonial galego e português* dedicado a textos escritos en gallego y en portugués.

Debido al gran número de corpus dedicados al español, el apartado que trata los corpus en esta lengua ha sido dividido en tres subapartados: Español europeo⁴, Español americano y Español europeo y americano.

Siguiendo la misma coherencia, el apartado dedicado al portugués también se debería dividir en subapartados puesto que también contiene corpus en portugués de distintos continentes. Sin embargo, debido al número reducido de corpus en esta lengua, de momento se agrupan todos en un solo apartado. Si se siguiera la lógica de separar por continentes, los cinco corpus portugueses (Galicia y Portugal (2); Brasil; Portugal + Brasil; Portugal, Brasil, Mozambique y Angola) se deberían subdividir en cuatro subapartados casi todos conteniendo un solo corpus (Portugués europeo; Portugués americano; Portugués europeo y americano; Portugués europeo, americano y africano), algo que ahora no parece muy razonable, pero que en el futuro, si crece el número de corpus en lengua portuguesa, se podrá hacer.

En cuanto al eje secundario en el que también se estructura *CORHIBER*, el del ámbito, hasta el momento se han establecido tres categorías:

- General.- Se refiere a los corpus que incluyen textos de todo tipo, literarios, documentales, religiosos, etc., como, por ejemplo, el *CIPM (Corpus Informatizado de Português Medieval)* o el *Gondomar (Corpus dixital de textos galegos da Idade Moderna)* que recogen textos de todos los tipos.
- Documental.- Concerniente a corpus que incluyen solamente textos de carácter documental, testamentos, cartas, fueros, etc., como el *CODEA+ 2015*, que solamente recoge documentos archivísticos, o el *CorLexin*, que únicamente incluye documentación notarial.
- Otros.- Cuando se trata de corpus que recogen textos especializados de carácter no documental, como, por ejemplo, el *Biblia Medieval*, que solamente ofrece versiones de la Biblia, o el *An Electronic Corpus of 15th Century Castilian Cancionero Manuscripts*, dedicado a los cancioneros.

4. INFORMACIONES QUE SE OFRECEN DE LOS RECURSOS PRESENTADOS EN *CORHIBER*

Las informaciones que *CORHIBER* ofrece de cada corpus que presenta se estructuran en cuatro apartados con distintos ítems en cada uno de ellos: A) Identificación y procedencia, B) Parámetros clasificatorios, C) Estado actual y D) Descripción.

A) Identificación y procedencia

En este apartado se da cuenta de los principales datos referidos a informaciones externas del corpus:

- Siglas del corpus
- Título del corpus

⁴ Inicialmente este subapartado se denominaba “Español peninsular”, sin embargo, Guillermo Rojo nos hizo notar que esta denominación excluía el español de las Islas Canarias.

- Dirección electrónica
- Centro o institución al que pertenece
- Nombre/s responsable/s
- Correo electrónico de contacto
- Año de inicio del corpus

B) Parámetros de clasificación

En este apartado se especifican los parámetros espaciotemporales y tipológicos en los que se define cada corpus. Se trata de los parámetros mínimos necesarios para poder situar las palabras de los textos que componen el corpus en su contexto «natural», puesto que estas se producen en un tiempo, en un espacio y en una situación comunicativa concretos y determinados.

- Ámbito temporal
- Ámbito geográfico
- Ámbito tipológico

C) Estado actual

Apartado en el que se describen aquellos aspectos del corpus reseñado que son susceptibles de que experimenten cambios a medida que avanza el proyecto.

- Estado de su consulta (En abierto / Cerrado / Por suscripción)
- Lematización (Sí / En proceso / No)
- Núm. de documentos
- Núm. de formas
- Fecha del recuento

D) Descripción

Este apartado ofrece una breve descripción de aquellos aspectos más relevantes del corpus. Dicha descripción ha sido facilitada por los propios directores de cada proyecto.

- Breve descripción del corpus

5. MANTENIMIENTO DEL PORTAL

El Portal *CORHIBER* se va a actualizar anualmente, ofreciendo una versión correspondiente a cada año. Sin embargo, ello no va a implicar que cuando se tenga noticia de una nueva aplicación y se haya comprobado que cumple los requisitos para

estar de *CORHIBER*⁵ se espere al final del año para darla a conocer, sino que el portal se actualizará constantemente, pero, al final de cada año se publicará la versión correspondiente.

De este modo, se pretende que el *Portal CORHIBER* sea una herramienta en constante evolución, lo que implica que debe de tener un mantenimiento y revisiones periódicas. Por ello, se actuará en dos frentes: A) Revisión y actualización de los datos de los corpus ya incorporados y B) Nuevas incorporaciones.

A) Revisión y actualización de los datos de los corpus ya incorporados

Para la revisión y actualización de los datos se seguirán dos vías:

- Al menos una vez al año se revisarán todos los enlaces que figuran en el portal para comprobar su vigencia y actualidad y, si es el caso, se suprimirán aquellos que hayan dejado de figurar en la Red.
- Así mismo, también una vez al año se escribirá a los directores de cada proyecto para actualizar los datos.

B) Nuevas incorporaciones

Para la obtención de nuevas incorporaciones de corpus que encajen dentro de los postulados de *CORHIBER*, se recurrirá a la revisión de los programas de congresos dedicados a la lingüística de corpus, especialmente el de *Corpus Diacrónicos en Lenguas Iberorrománicas (CoDiLI)*, así como a las noticias que podamos recibir a través de investigadores que trabajen en proyectos de corpus históricos iberorrománicos que nosotros desconozcamos.

Así, por ejemplo, en el *CoDiLI 5* se han dado noticias de distintos corpus que no figuran en *CORHIBER v. 2.0 (2020)* y que, se comprobará que existen en la Red y se escribirá a sus directores para, si es el caso, poder incluir-los en la versión de *CORHIBER v. 3.0 (2021)*.

- *CORDICan – Corpus documental de las Islas Canarias*
- *DIACOM*
- *CORPAT (Corpus de los Atlas Lingüísticos)*
- *Corpus de Documentos de Ávila de la Hispanic Society*
- *CIGCMod (Corpus Informatitzat de la Gramàtica del Català Modern)*
- *CoDiAJe (Corpus diacrónico anotado del judeoespañol)*
- *Corpus para la historia lingüística de la minoría morisca en las fronteras de Castilla (ss. XV-XVI)*
- *Oralia diacrónica del español: un nuevo corpus sobre el español de la Edad Moderna*

⁵ Los requisitos para figurar en *CORHIBER* son: tener todos los datos que figuran en la ficha de cada recurso, disponer de la descripción hecha por los responsables del recurso y que el recurso tenga una dirección activa en Internet.

Gracias a informaciones que nos han hecho llegar sus directores, en la versión de *CORHIBER v. 3.0 (2021)* se ha incorporado un nuevo apartado dedicado a los corpus Latino-romances.

6. ESTADO ACTUAL

La versión de *CORHIBER v. 3.0 (2021)*, incluye 39 corpus de contenido y formatos muy diversos. Por lenguas, el reparto es el siguiente:

- 3 Catalán
- 21 Español
- 4 Gallego
- 2 Judeoespañol
- 2 Latino-romance
- 3 Plurilingües
- 4 Portugués

Los 21 corpus del español se dividen en:

14 Español europeo:

7 de carácter documental (p. ej. *CODEMA Corpus diacrónico de documentación malagueña*).

5 de carácter general (p. ej. *Biblioteca Digital de Textos del Español Antiguo*) o de carácter específico (*Biblia Medieval*).

2 Español americano.

Todos de carácter general (p. ej. *Corpus Diacrónico y Diatópico del Español de América* y el *Corpus Electrónico del Español Colonial Mexicano*).

7 Español europeo y americano.

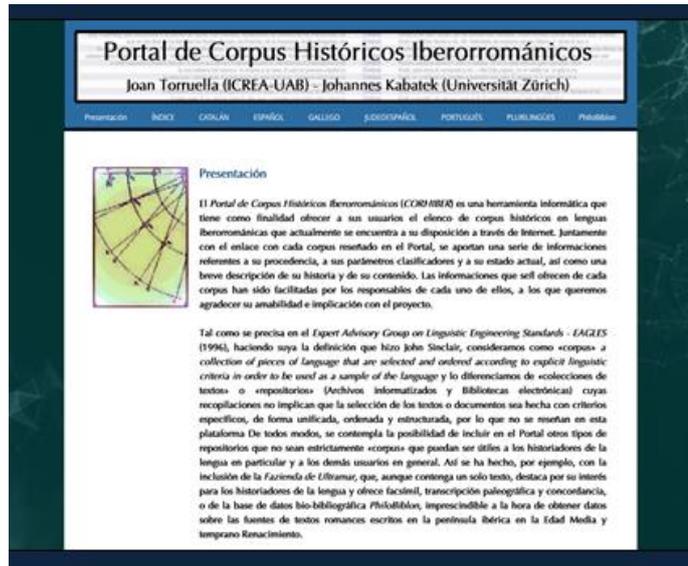
3 de carácter documental (p. ej. *CHARTA -Corpus Hispánico y Americano en la Red: Textos Antiguos*).

4 de carácter general (p. ej. *Corpus del Nuevo Diccionario Histórico del Español*).

Para la lista de todos los corpus presentes en *CORHIBER v.3.0 (2021)*, véase el apartado 8.

7. COMO FUNCIONA CORHIBER

Se accede a *CORHIBER* a partir de la dirección <http://corhiber.org>. La primera pantalla da acceso a una sucinta presentación y a la Barra de menús que se halla en la parte superior, justo debajo del título del Portal.



Portal de Corpus Históricos Iberorrománicos
Joan Torruella (ICREA-UAB) - Johannes Kabatek (Universität Zürich)

Presentación ÍNDICE CORUÁN ESPAÑOL GALLEGO JEROSOLIMITANO PORTUGUÉS PLURILINGÜES PIRIBIBES

Presentación

El Portal de Corpus Históricos Iberorrománicos (*CORHIBER*) es una herramienta informática que tiene como finalidad ofrecer a sus usuarios el elenco de corpus históricos en lenguas Iberorrománicas que actualmente se encuentra a su disposición a través de Internet. Juntamente con el enlace con cada corpus resuado en el Portal, se aportan una serie de informaciones referentes a su procedencia, a sus parámetros clasificadores y a su estado actual, así como una breve descripción de su historia y de su contenido. Las informaciones que señ ofrecen de cada corpus han sido facilitadas por los responsables de cada uno de ellos, a los que queremos agradecer su amabilidad e implicación con el proyecto.

Tal como se precisa en el *Expert Advisory Group on Linguistic Engineering Standards - EAGLES* (1996), haciendo suya la definición que hizo John Sinclair, consideramos como «corpus» a *collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language* y lo diferenciamos de «colecciones de textos» o «repositorios» (Archivos informatizados y Bibliotecas electrónicas) cuyas recopilaciones no implican que la selección de los textos o documentos sea hecha con criterios específicos, de forma unificada, ordenada y estructurada, por lo que no se montan en esta plataforma. De todos modos, se contempla la posibilidad de incluir en el Portal otros tipos de repositorios que no sean estrictamente «corpus» que puedan ser útiles a los historiadores de la lengua en particular y a los demás usuarios en general. Así se ha hecho, por ejemplo, con la inclusión de la *Fazenda de Ultramar*, que, aunque contenga un solo texto, destaca por su interés para los historiadores de la lengua y ofrece facsimil, transcripción paleográfica y concordancia, o de la base de datos bio-bibliográfica *Piribibes*, imprescindible a la hora de obtener datos sobre las fuentes de textos romances escritos en la península ibérica en la Edad Media y temprano Renacimiento.

A partir de la barra de menús se puede acceder al ÍNDICE de todos los recursos presentes en *CORHIBER*, en orden alfabético.



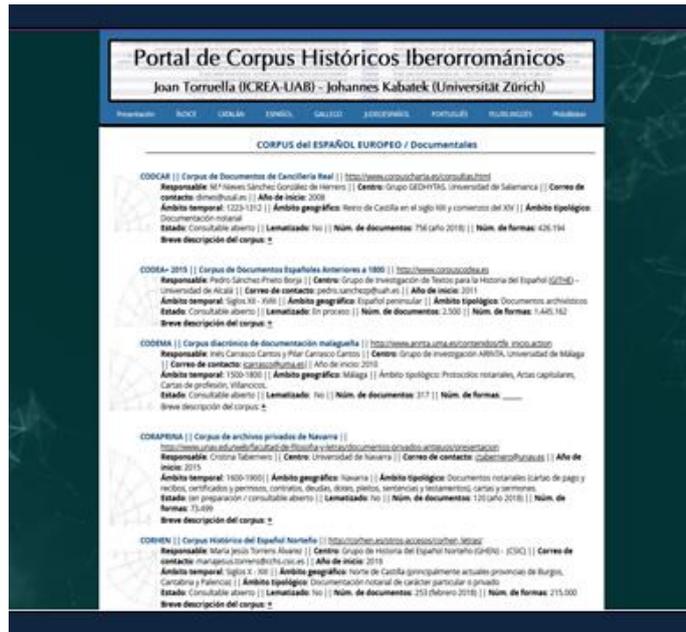
Portal de Corpus Históricos Iberorrománicos
Joan Torruella (ICREA-UAB) - Johannes Kabatek (Universität Zürich)

Presentación ÍNDICE CORUÁN ESPAÑOL GALLEGO JEROSOLIMITANO PORTUGUÉS PLURILINGÜES Piribibes

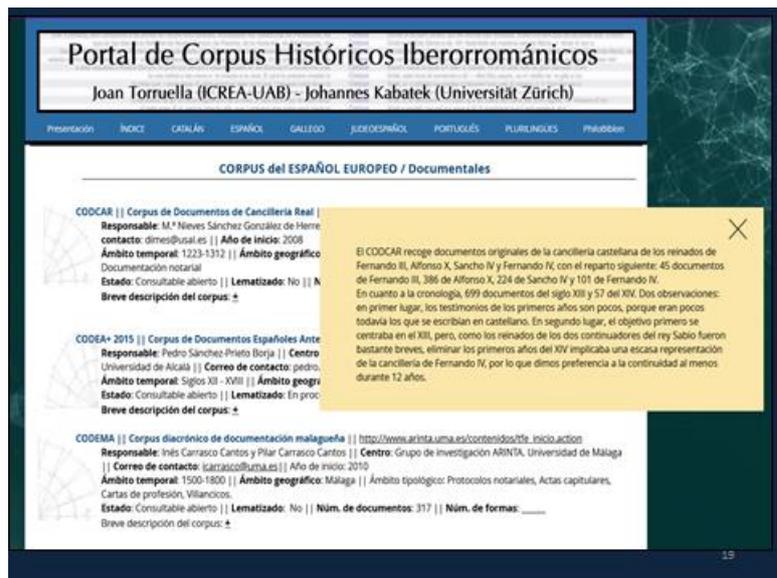
ÍNDICE DE CORPUS

- ADMOTE. I. I. Archivo Digital de Manuscritos y Textos Españoles
- An Electronic Corpus of 15th Century Castilian Cancionero Manuscripts
- BODIL. I. I. Diálogo Bodil
- Biblia Medieval
- BIFIA-DIGEST. I. I. Biblioteca Digital de Textos del Español Antigo / Digital Library of Old Spanish Texts
- CBH. I. I. Corpus del Nuevo Diccionario Histórico del Español
- CE. I. I. Corpus del Español
- CHARTA. I. I. Corpus Hispanico y Americano en la Red, Textos Antigos
- CHCA. I. I. Corpus Informático del Català Antic
- CHPI. I. I. Corpus Informático de Portugués Medieval
- CODICAR. I. I. Corpus de Documentos de Cancillería Real
- CODIMA. 2013. I. I. Corpus de Documentos Españoles Anteriores a 1800
- CODIMA. I. I. Corpus diacrónico de documentación malagueña
- CORAFINCA. I. I. Corpus de archivos privados de Navarra
- CORDE. I. I. Corpus Diacrónico del Español
- CORHIVAM. I. I. Corpus Diacrónico y Diatópico del Español de América
- CORHICOM. I. I. Corpus Diacrónico del Español Colonial Mexicano
- CORHIN. I. I. Corpus Histórico del Español Nortño
- CORHIS. I. I. Corpus Histórico Judío-español
- CorLentis. I. I. Corpus Lencin de los venecianos
- Corpus de portugués
- Corpus Histórico de portugués Tronco Brêves
- CORGALA. I. I. Corpus de Textos Antigos de Galicia
- CTA. I. I. CORPUS DE TEXTOS ANTIGOS em português até 1525
- DICCA XIV. I. I. Diccionario del castella del siglo XIV a la Corona d'Aragó
- DTICA. I. I. Diccionario de Textos Conocidos de Andalucía
- DTCA. I. I. Diccionario de Textos Catalans Antics
- En el uso del Interactio, Cartas de Ultramar a España
- Fazenda de ultramar
- GMR. I. I. Gallieses Monumenta Historica
- Gondomar. I. I. Gondomar, Corpus digital de textos galegos da Idade Moderna
- ODE. I. I. Online Diacrónico del español
- OSTA. I. I. Old Spanish Textual Archive
- P. S. I. Post Scriptum, A Digital Archive of Ordinary Writing
- Piribibes
- PIRIB. I. I. Projeto para a História do Português Brasileiro
- RIALC. I. I. Reserchers Informaticizzati della Antica Letteratura Catalana
- TALS. I. I. Tesouro Informaticizado da Língua Galega
- TLPSP. I. I. Tesouro da Língua petrimonial gallega e portuguesa

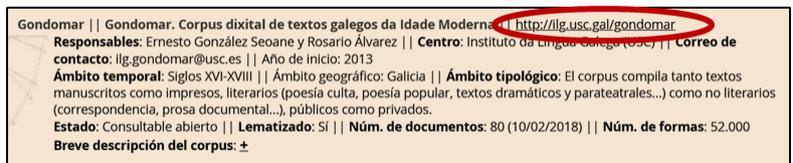
Haciendo clic en cualquiera de los recursos listados se accede a su ficha correspondiente. También se puede acceder a los distintos recursos a través de la lengua oportuna en la Barra de menús de la aplicación.



Haciendo clic en el signo + del apartado Descripción se accede a la presentación del corpus o del recurso seleccionado hecha por sus responsables.



Finalmente, haciendo clic en la dirección http que figura en segundo o tercer lugar de la ficha, se accede a la página principal del recurso seleccionado.



8. RECURSOS PRESENTES EN CORHIBER v. 3.0 (2021)

ADMYTE || *Archivo Digital de Manuscritos y Textos Españoles*
An Electronic Corpus of 15th Century Castilian Cancionero Manuscripts
BDDH || *Dialogyca BDDH*
Biblia Medieval
BIDTEA/DLOST || *Biblioteca Digital de Textos del Español Antiguo / Digital Library of Old Spanish Texts*
CDH || *Corpus del Nuevo Diccionario Histórico del Español*
CE || *Corpus del Español*
CHARTA || *Corpus Hispánico y Americano en la Red: Textos Antiguos*
CICA || *Corpus Informatizat del Català Antic*
CIPM || *Corpus Informatizado de Português Medieval*
CODCAR || *Corpus de Documentos de Cancillería Real*
CODEA+ 2015 || *Corpus de Documentos Españoles Anteriores a 1800*
CODEMA || *Corpus diacrónico de documentación malagueña*
CoDiAJe || *Corpus Diacrónico Anotado del Judeoespañol*
CODOLCAT || *Corpus Documentale Latinum Cataloniae*
CODOLGA || *Corpus documentale Latinum Gallaeciae*
CORAPRINA || *Corpus de archivos privados de Navarra*
CORDE || *Corpus Diacrónico del Español*
CORDIAM || *Corpus Diacrónico y Diatópico del Español de América*
COREECOM || *Corpus Electrónico del Español Colonial Mexicano*
CORHEN || *Corpus Histórico del Español Norteño*
CORHIJE || *Corpus Histórico Judeoespañol*
CorLexIn || *Corpus Léxico de Inventarios*
Corpus do português
Corpus Historico do Português Tycho Brahes
COTAGAL || *Corpus de Textos Antiguos de Galicia*
DiCCA XV || *Diccionari del castellà del segle XV a la Corona d'Aragó*
DITECA || *Diccionario de Textos Concejiles de Andalucía*
DTCA || *Diccionari de Textos Catalans Antics*
En el ojo del huracán. Cartas de Ultramar a España
Fazienda de ultramar
GMH || *Gallaeciae Monumenta Historica*
Gondomar || *Gondomar. Corpus dixital de textos galegos da Idade Moderna*
ODE || *Oralia diacrónica del español*
P. S. || *Post Scriptum. A Digital Archive of Ordinary Writting*
PhiloBiblon
PHPB || *Projecto para a História do Português Brasileiro*
RIALC || *Repertorio Informatizzato della'Antica Letteratura Catalana*
TILG || *Tesouro Informatizado da Lingua Galega*
TLPGP || *Tesouro do Léxico Patrimonial Galego e Português*