

Marta Punsola Munárriz - Pere J. Quetglas - Susanna Allés Torrent

LA EDICIÓN DIGITAL
DEL «GLOSSARIUM MEDIAE LATINITATIS CATALONIAE»
(GMLC)*

HISTORIA EDITORIAL DEL GMLC

El proyecto del *Glossarium Mediae Latinitatis Cataloniae* (GMLC) fue concebido e iniciado en la década de los cincuenta del siglo XX por el profesor M. Bassols de Climent y dirigido desde sus comienzos por el profesor J. Bastardas. Se trata de un proyecto de larga duración que, desde el año 1960, se viene publicando bajo el título *Glossarium Mediae Latinitatis Cataloniae, voces latinas y romances documentadas en fuentes catalanas del año 800 al 1100*. Las fuentes para la redacción del GMLC están constituidas fundamentalmente – aunque no exclusivamente – por documentación notarial redactada en lengua latina en los territorios del dominio lingüístico del catalán durante los siglos IX-XII. Este corpus documental se convierte en un testimonio privilegiado del latín medieval, especialmente de la lengua jurídica, eclesiástica e institucional de la Alta Edad Media; el latín de estos documentos presenta ya la influencia de la lengua romance hablada, al mismo tiempo que puede observarse en su redacción la existencia de lo que podríamos llamar un romance latinizado. Así pues, a partir de la confección de artículos lexicográficos que toman como punto de referencia dicha documentación, el GMLC se convierte en una herramienta imprescindible para filólogos, juristas, historiadores y, en general, para cualquier persona interesada en la Alta Edad Media catalana.

Con esta voluntad de crear una herramienta de consulta y ayuda, el equipo del GMLC desarrolla diversas líneas de investigación paralelas e íntimamente relacionadas entre sí. Por un lado, la redacción de artículos destinados a la publicación del glosario¹ y, por otro, la tarea de digitalización del corpus textual. Por lo que respecta a la redacción, en el año 2010 terminaron las tareas de preparación de una reedición del primer volumen (letras A-D), redactada en tres lenguas (catalán, castellano e inglés) y siguiendo unos criterios de edición uniformadores y que intentaban res-

Latinitad medieval hispánica. Edición de Juan Fco. Mesa Sanz, Firenze, SISMEL - Edizioni del Galluzzo, 2016

ponder a unas líneas directrices claras². En referencia a la tarea de informatización del corpus, desde el año 2000 se han ido digitalizando más de 22.000 documentos, los cuales constituyen un corpus digital muy útil para el vaciado y la corrección de los textos. El equipo del GMLC, fiel a la voluntad de crear una herramienta de ayuda tanto para los investigadores como para el público interesado en general, empezó a confeccionar en el año 2010 una base de datos léxica destinada a hacer accesible a través de internet este vasto corpus digital hasta el momento de uso exclusivamente interno. Así, en mayo de 2012 salió a la luz la primera versión del *Corpus Documentale Latinum Cataloniae* (CODOLCAT), base de datos accesible en línea (<http://gmlc.imf.csic.es/codolcat>) que ofrece al usuario la posibilidad de realizar búsquedas léxicas en el corpus textual del GMLC. El CODOLCAT está concebido como una publicación periódica que se actualiza anualmente, y que consta, en la versión de 2013 (v. 2), de más de 2000 documentos introducidos.

Como se desprende de lo expuesto hasta ahora, en los últimos años el GMLC se ha mantenido fiel a la tarea lexicográfica originaria de redacción de artículos, al mismo tiempo que aprovecha los recientes avances tecnológicos con el fin de optimizar la difusión y la accesibilidad de su rico corpus documental.

Llegados a este punto, y con la reedición del primer volumen sin publicar en papel, el grupo se planteó la posibilidad de unir sus dos vertientes, la redacción lexicográfica tradicional y la inmediatez proporcionada por una herramienta en línea como el CODOLCAT. En consecuencia, y como resultado de la evolución del grupo a lo largo de los años, la dirección del GMLC toma conciencia de la importancia de la aplicación de las nuevas tecnologías al estudio de las humanidades y, en consonancia con otros grupos europeos dedicados a la edición de obras lexicográficas de gran envergadura, surge el proyecto de llevar a cabo una edición en versión digital que sea consultable en línea de manera gratuita. Esta versión digital toma como punto de partida la reedición trilingüe del primer volumen correspondiente a las letras A-D, empezando como programa piloto con la publicación electrónica de las voces correspondientes a la letra A.

Paralelamente a la redacción de nuevas voces, a la ampliación de la base de datos del corpus textual (CODOLCAT) y al diseño y elaboración de la edición digital del GMLC, en el seno del grupo se realizan diversas investigaciones relativas a la documentación latina medieval catalana, del mismo modo que se están elaborando varias tesis doctorales que abordan distintos aspectos de la lengua de dicha documentación.

EDICIÓN DIGITAL DEL GMLC

En las próximas páginas, nos centraremos concretamente en esta última fase de elaboración de la edición digital. Para hablar de la versión electrónica del GMLC, resulta pertinente destacar, en primer lugar, algunas de las ventajas que supone optar por un marcado digital frente a una edición en formato papel. Una vez subrayadas estas ventajas principales, procederemos a explicar qué tipo de marcado ha decidido utilizar el grupo del GMLC, así como qué procedimiento de trabajo se ha seguido.

VENTAJAS DEL MARCADO DIGITAL

Como decíamos, pues, la edición en formato electrónico supone una serie de ventajas respecto a la edición en papel, las cuales conviene subrayar como punto de partida.

En primer lugar, cabe destacar la posibilidad de ofrecer un acceso permanente y gratuito a través de la web, lo cual facilita la consulta del GMLC no sólo a un público especializado, sino también a uno más amplio. Como ya se ha dicho, el equipo se rige por la voluntad de crear una herramienta de trabajo y de consulta que sea útil y de fácil acceso; en este sentido, una edición en línea se adecua perfectamente a esta voluntad, otorgándole al GMLC una mayor difusión.

Por otro lado, el marcado digital multiplica sustancialmente las posibilidades de presentación del contenido, permitiendo superar las restricciones tradicionales que impone el formato en papel. Así, la distribución de cada una de las entradas del glosario puede ser modificada en función de las necesidades o de los intereses tanto de los editores como de los usuarios. Cada entrada, por ejemplo, se puede distribuir jerárquicamente basándose en diversos tonos de color que indiquen los distintos elementos de la voz, a saber, variantes ortográficas, etimología, construcciones morfológicas, etc.

Además de enriquecer la distribución y presentación del contenido textual del GMLC, la versión digital es más ágil en lo que a la búsqueda de términos se refiere. Y es que, aparte de permitir la búsqueda sencilla de voces, ofrecerá la posibilidad de realizar búsquedas avanzadas a partir de una serie de parámetros como son la cronología, la etimología o la categoría gramatical, tal y como ya veremos un poco más adelante.

Cabe destacar también el hecho de que gracias al sistema de marcado utilizado se podrá explotar información textual contenida en el glosario.

MARTA PUNSOLA MUNÁRRIZ - PERE J. QUETGLAS - SUSANNA ALLÉS TORRENT

Así, por ejemplo, se podrán activar remisiones internas, crear enlaces directos con otras obras citadas, recuperar las referencias bibliográficas de manera inmediata, desarrollar las abreviaturas, etc.

Finalmente, nos parece importante destacar que la edición en formato digital facilita notablemente la corrección del texto, así como la rapidez y agilidad en la reedición.

TIPO DE MARCADO: «TEXT ENCODING INITIATIVE» (TEI)

Para la confección de la versión digital, el equipo del GMLC ha optado por una informatización en lenguaje XML (*eXtensible Markup Language*), siguiendo las recomendaciones o directrices de la *Text Encoding Initiative* (TEI). TEI es un estándar de codificación especialmente apto para el procesamiento y la presentación en formato digital de textos en humanidades y ciencias sociales y, en sus recomendaciones, se define un modelo de marcado general para las fuentes lexicográficas. Así pues, el XML-TEI constituye una opción adecuada a la hora de crear la versión en formato electrónico de una obra lexicográfica compleja como es el GMLC.

Cabe destacar, por otro lado, que la elección de este sistema se inspira en otros proyectos europeos que están desarrollando una labor de informatización de sus respectivos diccionarios basándose en este mismo tipo de marcado. El caso más notable lo constituye la versión en línea del *Du Cange* (<http://ducange.enc.sorbonne.fr/>), realizada en el seno de la *École Nationale des Chartes*; a este proyecto cabe sumar las tareas de informatización realizadas por el equipo de redacción del *Nouum Glossarium Mediae Latinitatis* (NGML) (<http://www.glossaria.eu/ngml/>) en el *Institut de Recherche et d'Histoire des Textes* (CNRS) de París. Del mismo modo, destacan los proyectos de digitalización del *eLexicon Mediae Latinitatis Polonorum* (<http://scriptores.pl/lexicon/>), con sede en el *Instytut Języka Polskiego* (*Polska Akademia Nauk*) de Cracovia, del *Diccionario griego-español* (<http://dge.cchs.csic.es/xdge/>), realizado en el Centro de Ciencias Humanas y Sociales del CSIC (Madrid), o del *Mittelateinisches Wörterbuch* (http://www.woerterbuch-portal.de/woebus_alle/Woebu6?krit), en la *Bayerische Akademie der Wissenschaften* en Múnich. Precisamente con la voluntad de interlocución y de puesta en común de los métodos y soluciones adoptados por estos y otros proyectos, el pasado mes de mayo de 2013 se celebró en Barcelona un seminario bajo el título *XML-TEI para obras lexicográficas de la Antigüedad y la Edad Media* (<http://gmlc.imf.csic.es/2013/Workshop/>), que reunió tanto a miembros de los equipos ya mencionados, como a especialistas internacionales en el uso de este sistema de

L Języka

marcado aplicado a la digitalización de obras en humanidades y ciencias sociales.

«WORKFLOW» O FLUJO DE TRABAJO

Una vez determinado el sistema de marcado elegido por el grupo, pasamos a continuación a describir brevemente el proceso de trabajo que se ha seguido y las distintas fases por las que ha pasado la creación de la versión digital del GMLC.

En primer lugar, a partir de un documento Word con el texto de la reedición no publicada de la letra A, se ha realizado una conversión automática en formato .xml³. Una vez determinada la estructura de las voces, se ha creado un esquema RELAXNG⁴ destinado a estructurar jerárquicamente el documento .xml, y se han redactado unas instrucciones de marcado de uso interno también en formato digital, las cuales se han ido ampliando y perfilando a lo largo del proceso de marcado. Así, paulatinamente, se ha modificado y adaptado el fichero .xml a un marcado filológico y lexicográfico del texto, atendiendo en todo momento a las posibles necesidades concretas de los futuros usuarios.

El marcado de la letra A ha sido realizado fundamentalmente por la doctora Susanna Allés, coordinadora de la edición digital y máster en “Technologies numériques appliquées à l’histoire” por la *École Nationale des Chartes*, y Frédérique Laugrost, estudiante del mismo máster que realizó, en 2013, una estancia de tres meses en la sede del GMLC. También han colaborado en el proceso de marcado Marta Segarrés y Marta Punsola, las cuales se encargan paralelamente de alimentar la base de datos léxica CODOLCAT. El trabajo se ha desarrollado bajo la supervisión de la doctora Ana Gómez Rabal y de Pere Farró Sales, quien se ha encargado de revisar los estándares de construcción y de crear el motor de búsqueda y la interfaz web que van a permitir llegar a la fase final de publicación en línea del GMLC.

EVOLUCIÓN DEL GMLC

Para ilustrar la evolución que ha experimentado el GMLC a lo largo de los años, les ofrecemos un recorrido en imágenes que ilustran el paso de una ficha originaria a la versión en papel (se trata, en concreto, de una imagen de la reedición del primer volumen) y cómo, a partir del marcado de ésta, se llega a la fase final de visualización de la versión digital.

MARTA PUNSOLA MUNÁRRIZ - PERE J. QUETGLAS - SUSANNA ALLÉS TORRENT

abatas *abbata* **sig.**

**abatas vero et monachas Deo servientes
in domum sancti Iohannis**

927
doc. San Juan Abadesas, a. A. C. A.
ed. Udina, "E. Arch. Condal de Barñs. s. IX X", pág. 225 doc. 86

abatas = abbatissae

I. Martínez

Ficha de la voz *abbata*

abbata (aba-) [*ab* *abbatus*, cf. s. v. *abbas*] *abadessa* | *abadesa* | *abbess*: 927
(orig.?, S. Joan de les Abadeses) ACondal 86, p. 225: abatas uero et monachas Deo
servientes in domum sancti Iohannis ... obtineant ipsa omnia quod superius insertum
est.

La voz *abbata* en la reedición del primer volumen

```
<entry xml:id="abbata">
  <form type="lemma">
    <orth n="1">abbata</orth> (<orth n="2" expand="abata">aba-</orth>)</form>
    <gramGrp>
      <pos value="subst"/>
    </gramGrp>
    <etym><abbr>ab</abbr>
      <term xml:lang="la">abbatus</term>, <abbr>cf.</abbr>
      <abbr>s. v.</abbr>
      <ref target="#abbas">abbas</ref></etym>
    <sense level="1"><def xml:lang="ca">abadessa</def>
      <def xml:lang="es">abadesa</def>
      <def xml:lang="en">abbess</def>: <cit type="exemplum">
        <bibl corresp="#ACondal"><date when="9927">927</date> (orig.?, S. Joan
        de les Abadeses) ACondal <biblScope>86, p. 225</biblScope></bibl>
        <quote xml:lang="la">abatas uero et monachas Deo servientes in domum
        sancti Iohannis ... obtineant ipsa omnia quod superius insertum
        est.</quote>
      </cit></sense>
  </entry>
```

Marcado XML-TEI de la voz *abbata***abbata** (aba-)[*ab* *abbatus*, cf. s. v. *abbas*]*abadessa* | *abadesa* | *abbess*:

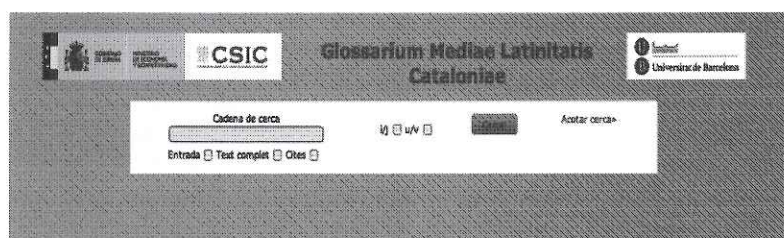
927 (orig.?, S. Joan de les Abadeses) ACondal 86, p. 225: abatas uero et monachas Deo servientes in domum sancti Iohannis ... obtineant ipsa omnia quod superius insertum est.

La voz *abbata* en la versión digital

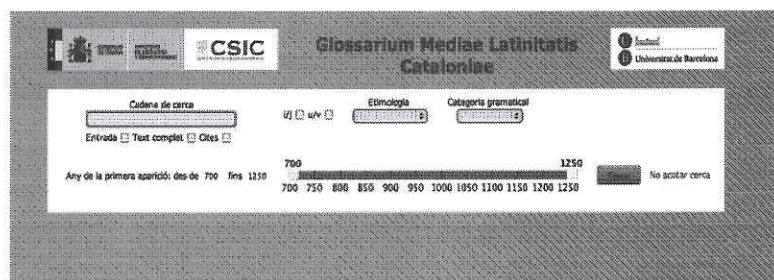
LA EDICIÓN DIGITAL DEL «GLOSSARIUM MEDIAE LATINITATIS CATALONIAE» (GMLC)

MUESTRA DE LA EDICIÓN DIGITAL

Para llegar a esta visualización final de las voces, los usuarios del GMLC en línea van a disponer de un buscador que les va a permitir realizar diversos tipos de búsqueda. Por un lado, se podrán llevar a cabo búsquedas simples, al mismo tiempo que se ofrecerá la posibilidad de escoger entre una búsqueda por entradas, por texto completo o por citas.



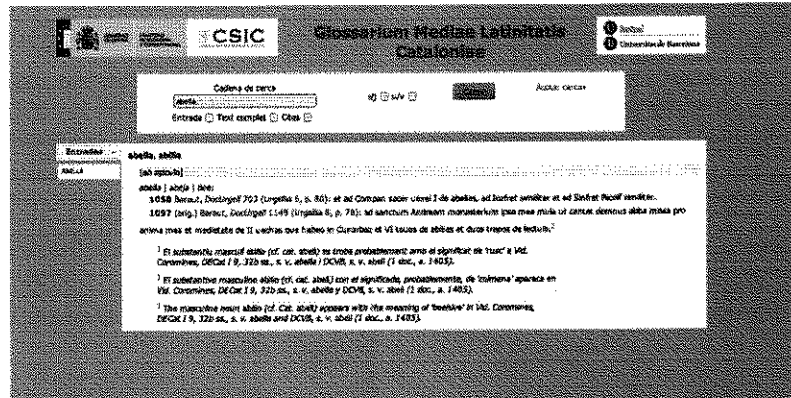
Por otro lado, también se podrán realizar búsquedas avanzadas en función de distintos parámetros. Así, será posible filtrar las búsquedas por etimología, por categoría gramatical o por cronología.



Tanto la búsqueda sencilla de términos como la avanzada van a permitir diversas opciones de introducción del lema que se desee consultar. En este sentido, se podrán introducir palabras completas, voces truncadas, secuencias, etc., en base a unas normas de introducción similares a las ya presentes en el CODOLCAT.

Finalmente, esta será la visualización del resultado de la búsqueda, aunque se debe destacar que lo que muestra la imagen es una versión beta de la edición, es decir, que no se trata todavía de la versión definitiva que se va a publicar en línea.

MARTA PUNSOLA MUNÁRRIZ - PERE J. QUETGLAS - SUSANNA ALLÉS TORRENT



PERSPECTIVAS DE FUTURO

Para concluir, querríamos apuntar algunas de las perspectivas de futuro del grupo, como son, por ejemplo, la conexión de la edición digital del GMLC con el CODOLCAT.

Por otro lado, esperamos poder ofrecer cada una de las entradas también en formato .pdf, para que el usuario tenga a su disposición una visualización más próxima a la versión en papel, satisfaciendo así las necesidades y demandas de todos los públicos a los cuales se dirige la edición electrónica de nuestra obra lexicográfica.

En otro sentido, la intención del equipo del GMLC es llegar a conectar la edición digital con otros diccionarios u obras de referencia en línea, como por ejemplo el Du Cange o el *Diccionario griego-español*, de manera que cada vez que alguna de estas obras sea citada en el GMLC se pueda acceder al portal web correspondiente a través de un solo *click*.

Finalmente, y como es lógico, el principal objetivo a largo plazo del equipo es poder ofrecer la publicación en línea íntegra del GMLC, tanto de los fascículos ya publicados en versión papel como de las letras restantes que todavía no han salido a la luz. Así, y paralelamente a la labor de alimentación del CODOLCAT, el grupo se reafirma en su voluntad de ofrecer una herramienta de consulta fácilmente accesible a todo el público interesado que contribuya, en última instancia, a un mejor conocimiento de la Alta Edad Media en los territorios del dominio lingüístico del catalán.

NOTE

* Este artículo se ha realizado en el seno del equipo del *Glossarium Mediae Latinitatis Cataloniae*, beneficiario del proyecto "Informatización del *Glossarium Mediae Latinitatis Cataloniae* (7)" (FFI2012-38077-C02-00), subproyectos "Ampliación y desarrollo de la base de datos *Corpus Documentale Latinum Cataloniae* (CODOLCAT)" (FFI2012-38077-C02-01) y "Redacción del *Glossarium Mediae Latinitatis Cataloniae* (GMLC)" (FFI2012-38077-C02-02), sufragados por el Ministerio de Economía y Competitividad, y beneficiario, asimismo, del reconocimiento y la financiación como Grupo de Investigación Consolidado (2009SGR705) de la Generalitat de Cataluña.

1. Hasta ahora se han publicado los volúmenes correspondientes a las letras A-D, F y G (*Glossarium Mediae Latinitatis Cataloniae* conditum ab M. Bassols de Climent, conficiendum curavit Iohannes Bastardas, vol. I (A-D), Barcinone 1960-1985; J. BASTARDAS (dir.), *Glossarium Mediae Latinitatis Cataloniae*, fasc. 11 (F) y fasc. 12 (G), Barcelona: CSIC, 2001 y 2006, respectivamente). El equipo se encuentra actualmente elaborando la redacción de la letra E.

2. Sobre la preparación de la reedición del primer volumen, vid. A. Gómez Rabal, "En torno a la nueva edición del *Glossarium Mediae Latinitatis Cataloniae*" en J. Martínez Gázquez, Ó. de la Cruz Palma, C. Ferrero Hernández (eds.): *Estudios de latín medieval hispánico*, Firenze, SISMELE - Edizioni del Galluzzo, 2011, pp. 1075-1082 (Millenio Medievale, 92; Strumenti e Studi n. s., 30). La reedición en papel de la obra, sin embargo, no ha salido a la luz.

3. Esta primera transformación se ha realizado a través de OxGarage Conversion, aplicación que ofrece la propia plataforma TEI (<http://www.tei-c.org/oxgarage/#>).

4. RELAXNG (*REGular LAnguage for XML Next Generation*) es un lenguaje estándar que permite la creación de un esquema y una sintaxis específica sobre la que se basa el documento XML (equivale a las DTD, heredadas del SGML pero, a diferencia de estas, está escrita en XML). La especificación puede encontrarse en: <http://relaxng.org/>. Estos esquemas son de fácil creación gracias a la plataforma ROMA, también puesta a disposición por el TEI Consortium (<http://www.tei-c.org/Roma/>).

L (2014SGR929)