

Análisis semántico a partir de una base de datos relacional SQL: un ejemplo práctico

Dirección

Clara Martínez
Cantón

Gimena del Río
Riande

Francisco Barrón

Semantic Analysis from a Relational SQL Database: A Practical Example

Álvaro PIQUERO

Universidad Complutense de Madrid

alvaropiquero@ucm.es

<https://orcid.org/0000-0002-8352-3493>

Editora asociada

Romina De León

RESUMEN

La tesis doctoral *La imaginería en la poesía erótica de los Siglos de Oro* se centró en el análisis del vocabulario de más de quinientos poemas eróticos de los Siglos de Oro. Con cerca de 1.150 palabras analizadas a lo largo del estudio, la redacción del trabajo requirió el desarrollo de una metodología de base de datos relacional SQL para organizar, recuperar e interpretar la información de una forma estructurada y clara. El objetivo de esta nota de investigación es describir detalladamente la creación, estructura y la aplicación práctica de este método de investigación, además de demostrar que las bases de datos relacionales SQL pueden ser una herramienta adecuada para el análisis léxico-semántico de textos literarios. En resumen, este trabajo pretende ofrecer a la comunidad científica las bases teóricas para el desarrollo de proyectos similares teórica y metodológicamente.

ABSTRACT

The PhD project *La imaginería en la poesía erótica de los Siglos de Oro* (Imagery in the erotic poetry of Golden Age) was focused on the analysis of the vocabulary of more than five hundred Spanish sexual poems from the 16th and 17th centuries. With more than 1,150 words annotated throughout the study, the writing of the work required the development of a digital SQL relational database methodology to organise and retrieve the information in an orderly structure. The purpose of this contribution is to describe in detail its creation, its structure, and its practical application in research, apart from demonstrate that SQL relational databases can be a suitable tool for the lexical-semantic analysis of literary texts. To sum up, this work aims to offer other researchers some theoretical basis for the development of similar projects in terms of theory or methodology.

PALABRAS CLAVE

base de datos, SQL, semántica, erotismo, vocabulario.

KEYWORDS

Database, SQL, Semantics, Eroticism, Vocabulary.

RHD 8 (2023)

ISSN

2531-1786



La intención de este breve trabajo es describir el proceso de creación y desarrollo de una metodología digital para la redacción de la tesis doctoral *La imaginería en la poesía erótica de los Siglos de Oro* (Piquero, 2021), cuya intención fundamental fue abordar el estudio cualitativo y cuantitativo del mayor número posible de vocabulario sexual extraído de la poesía erótica de los siglos XVI y XVII.

1. PLANTEAMIENTO TEÓRICO Y OBJETIVOS

Desde un punto de vista filológico, para abordar un análisis pormenorizado del vocabulario sexual en la poesía erótica es primordial formar un corpus textual lo más amplio y representativo posible. En el caso del estudio arriba citado, los criterios de selección fueron fundamentalmente dos: en primer lugar, textos que utilizan un lenguaje sexual explícito y unívoco – grupo de control–; en segundo, textos que aluden al sexo con un vocabulario equívoco y ambiguo –grupo de experimentación–.

Además de estas fuentes primarias, se atendió a la mayor cantidad de bibliografía secundaria posible, desde la lingüística y la literatura hasta la historiografía, la etnografía o el folclore, entre otras materias. Con ello, se pretendía ofrecer al lector información sobre el contexto cultural y extralingüístico en el que se escribieron los poemas, que aclara y complementa el análisis puramente textual¹.

La teoría planteada arriba se apoya en las técnicas de interpretación literaria habituales – y sobradamente probadas–, como la comparación, la deducción o la intertextualidad, y no ofrece ningún escollo metodológico en un corpus textual reducido. El objetivo principal del trabajo, sin embargo, era abordar la lectura, comparación e interpretación de la mayor cantidad de textos e imágenes eróticas posible –en concreto, 549 poemas y 1130 imágenes²–. Ante esta ingente cantidad de información, la única posibilidad de desarrollar la tarea exitosamente es la aplicación de técnicas de Humanidades Digitales.

Hoy existen pocas dudas de que la codificación y el procesamiento de los textos literarios permite obtener una nueva dimensión analítica sobre ellos y abre la puerta a puntos de vista inéditos hasta ahora³. Como explicaba Dino Buzzeti hace ya más de diez años (2009, p. 46):

“image”, or representation of the text in digital form, can considerably enhance our opportunities of penetrating deeply into its discourse [...] In my opinion, the true rationale of a genuine digital edition consists precisely in taking advantage of the digital form of

¹ Estos tres puntos básicos –corpus de control, corpus hipotético y materiales complementarios– resumen las ideas vertidas por Whinnom (1981, p. 34), Vasvári (1988, p. 2), Débax (1989, p. 44), Lacarra Lanz (1996, p. 420), Díez Fernández (2019, p. 64) y, sobre todo, Garrote Bernal (2012, pp. 250-251; revisado en 2020, pp. 137-140), que han tratado de ofrecer una base teórica a la interpretación y análisis de los textos eróticos.

² Tras el proceso de revisión y ampliación al que se está sometiendo el trabajo para su publicación, la cantidad de textos ha crecido hasta los 660; no obstante, las cifras que se ofrecerán a lo largo de este artículo se corresponden con las que se utilizaron durante la redacción de la tesis doctoral.

³ Sobre el objetivo de las Humanidades Digitales en la actualidad, véase, entre otros, McGann (2004, pp. 75-124); Spence (2014, pp. 118-131 y 2015, pp. 8-12); y Allés-Torrent (2015, p. 18; 2017, p. 21 y 2021).

representation to improve our critical engagement with the text through effective computational processing⁴.

Aunque todavía nos encontramos inmersos en una casuística y una metodología inestables (Buzzeti, 2009; Fradejas Rueda, 2009-2010; McGann, 2014, pp. 125-198; Revenga, 2014; Spence, 2014 y 2015; Allés-Torrent, 2015, 2017 y 2021)⁵, la comunidad científica camina firmemente hacia la regularización metodológica, especialmente en lo que respecta al desarrollo de estudios basados en estándares digitales como la estilometría y el marcado XML-TEI. En base a ello, los primeros pasos del estudio estuvieron enfocados a plantear el método más adecuado para el análisis semántico del vocabulario sexual⁶.

Comenzando por el primero de los citados, la aplicación de análisis estilométricos a la literatura erótica, cuyo corpus adolece históricamente de atribuciones poco fiables, ha ofrecido nuevas evidencias para adscribir casi definitivamente el *Jardín de Venus* a la pluma de fray Melchor de la Serna (Blasco, 2015); y, en el ámbito concreto del léxico, se ha utilizado para mostrar los “dos universos verbales (y por extensión, morales) perfectamente diferenciados” (Blasco, 2020, p. 24) que habría entre la tradición petrarquista y la antipetrarquista –esto es, la erótica– (Blasco, 2020, pp. 22-27)⁷.

Fuera de estos dos acercamientos al corpus erótico, esta metodología no ofrecería datos relevantes a la hora de interpretar las subjetivas disemias en las que se apoya el vocabulario sexual, por lo que fue rápidamente descartada para el proyecto.

Una opción más realista para las pretensiones del trabajo la representaba el marcado XML-TEI (*Text Encoding Initiative*), que puede considerarse hoy el estándar filológico digital para el marcado de textos (Allés-Torrent, 2015, p. 18)⁸.

Sin contar los esfuerzos puramente editoriales, al que están dedicados la mayor parte de los proyectos⁹, es sumamente interesante el etiquetado semántico que algunos investigadores han

⁴ Precisamente por esta “forma digital” Susanna Allés-Torrent (2017) defiende que “una edición digital no se puede imprimir sin una pérdida de información y/o de funcionalidad (es decir, si lo que obtenemos al imprimir es el mismo resultado que las opciones que tenemos en pantalla, no es una edición académica digital) (...) la edición digital está guiada por un paradigma diferente al espacio bidimensional de la ‘página’ impresa y a los medios tipográficos de representación de la información” (p. 21).

⁵ Salvo Buzzetti y McGann, que buscan una perspectiva más general, se citan aquí fundamentalmente aproximaciones a las Humanidades Digitales centradas en la literatura hispánica.

⁶ Dada la intención descriptiva de este trabajo, en los párrafos siguientes se presentarán exclusivamente las opciones metodológicas que se tuvieron en cuenta para la preparación de la tesis doctoral, que no agotan las posibilidades que ofrece la filología digital en la actualidad.

⁷ Los resultados obtenidos a partir del análisis con la herramienta Voyant Tools, presentados en tres tablas y una gráfica, son verdaderamente elocuentes desde un punto de vista semántico general –*distanc reading*–. Antes de abordar el análisis específico del léxico, el investigador utiliza también LIWC para caracterizar la lengua de los textos. Aunque estos resultados son también interesantes, a mi juicio, son mucho más representativos los que consigue en el análisis de palabras.

⁸ Más allá de las Guidelines, <https://tei-c.org/guidelines/>, una amplia descripción de las características técnicas de lenguaje de marcado XML-TEI y sus posibilidades de marcado puede encontrarse en Buzzetti (2009, pp. 49-61), Allés-Torrent (2015, pp. 18-20) o Spence (2014, p. 121).

⁹ Para una revisión del estado actual de las ediciones XML-TEI en la literatura áurea, tanto internacional como nacional, véanse los estudios de Allés-Torrent (2015 y 2017). Quiero resaltar especialmente en este sentido el proyecto Garrett Online (<https://garrettonline.romanceiro.pt/en/>), cuya plataforma se encuentra todavía en construcción. Dirigido por la profesora Sandra Boto, el proyecto se centra en la edición de los romances del autor portugués y sus primeros resultados ofrecen resultados prometedores en cuanto a las posibilidades de edición, marcación y visualización de textos (<https://garrettonline.romanceiro.pt/>

ensayado en corpus textuales más amplios¹⁰.

Un primer ejemplo sería ADSO (Análisis distante del soneto castellano de los Siglos de Oro) ubicado en la Universidad de Alicante¹¹. Como explica Borja Navarro-Colorado (2015, pp. 105-113), el propósito de su proyecto fue desarrollar una codificación TEI de más de 5.000 sonetos en dos fases: en primer lugar, el marcado métrico a través de una herramienta de escansión automática; en segundo lugar, la aplicación de modelos de semántica distribucional para estudiar la regularidad semántica de las composiciones y agrupar poetas y sonetos en relación a estos significados.

Más allá de los resultados provisionales que todavía hoy ofrece su web¹², la aplicación de herramientas creadas a partir lenguajes de programación como Java –MALLET (Machine Learning for Language Toolkit)– o Python –Pycluster– permitieron al autor demostrar a partir de una lectura automática objetiva de los sonetos que la clásica separación temática, estilística y métrica entre los poetas del Renacimiento y Barroco sigue siendo perfectamente válida (Navarro-Colorado, 2015, pp. 108-112). De hecho, los resultados son tan específicos que llega a concluir, por ejemplo, que el tema funerario es infinitamente más recurrente en los poetas barrocos, sobre todo en Francisco de Quevedo (p. 109), o que cada tópico literario parece tener un ritmo métrico concreto asociado (p. 111)¹³.

Otros dos proyectos de investigación pioneros en España a la hora de aplicar un etiquetado XML-TEI a un corpus textual extenso han sido Bibliografía de Escritorias Españolas (BIESES)¹⁴, ubicado en la Universidad Nacional de Educación a Distancia (UNED), y Sujeto e Institución Literaria en la Edad Moderna (SILEM)¹⁵, que se coordina entre la Universidad de Córdoba y la de Sevilla y que, en la línea de colaboración que demandan las Humanidades Digitales, partió de los avances del primero para desarrollar su propio esquema de etiquetado¹⁶.

A diferencia ADSO, donde primaba la lectura automática y el *distant reading*, en estos proyectos lo que se pretendía era una edición académica TEI-XML por parte de los propios investigadores a partir de un esquema de datos enriquecido en el que, además de las marcas puramente estructurales, se debía incluir una amplia ontología de etiquetas conceptuales que permitieran recuperar información semántica sobre los textos y sus autores.

Lógicamente, en este caso el paradigma metodológico se acercaría más al del *close reading*, ya que son los propios investigadores quienes, partiendo de un esquema de datos

romanceiro/livro-i/7-o-anjo-e-a-princesa/).

¹⁰ Esta breve descripción de algunos proyectos y metodologías de diversa procedencia e índole no pretende ser exhaustiva y representa únicamente las opciones que se tuvieron en cuenta a la hora de plantear la aplicación de XML-TEI al trabajo.

¹¹ Accesible desde: <http://adso.gpsi.es/index.php/es/proyecto-adso>.

¹² Accesible desde: <http://goldenage.cervantesvirtual.com/>.

¹³ Aunque, según el propio investigador, en este último caso habría que recurrir al *close reading* para confirmar la hipótesis que se puede deducir del *distant reading* realizado por la computadora.

¹⁴ Accesible desde: <http://www.bieses.net/>.

¹⁵ Accesible desde: <http://www.uco.es/investigacion/proyectos/silem/index.php>.

¹⁶ Yo mismo tuve la oportunidad de trabajar como técnico en ambos proyectos, de manera que en los siguientes párrafos primará el conocimiento personal de su metodología sobre las posibles lecturas bibliográficas.

conjunto, deciden qué etiquetas deben aparecer en un texto y dónde han de incluirse para que la información que el usuario recopile después sea relevante.

Como resultado de esta codificación textual, ambos proyectos ofrecen un avanzado motor de búsqueda en el que se puede consultar la edición TEI completa –registrada por autor o siglo– o realizar búsquedas semánticas específicas a partir de las etiquetas disponibles, que reflejan cuestiones referidas a la obra –genéricas, estilísticas, métricas, etc.–, al autor o autora –valores, estatus, trayectoria, (auto)representación, etc.– o al contexto editorial –mediadores, familia, juicios críticos, etc.–¹⁷.

Los resultados, con sus diferentes aplicaciones metodológicas, son a mi juicio muy interesantes, puesto que consiguen uno de los objetivos principales de la filología digital: revisar y releer los textos desde nuevas perspectivas y ofrecer al investigador datos, lecturas y enfoques inalcanzables con métodos tradicionales.

Habida cuenta de la similitud que existe entre estas dos propuestas y el análisis que se pretendía desarrollar en la tesis citada arriba, en un primer momento se valoró muy positivamente la opción de realizar un etiquetado TEI del corpus de la poesía erótica áurea para recuperar posteriormente toda la información referida a la imaginaria sexual. Los escollos que esta técnica presentaba para un estudio individual tan amplio resultaron, sin embargo, insalvables.

En primer lugar, a pesar del extraordinario esfuerzo del proyecto Eros & Logos, no existe aún un corpus fiable digital de poesía erótica áurea, por lo que sigue siendo necesario recurrir a las ediciones físicas de los textos¹⁸. Dado que el corpus del trabajo estaba conformado por 549 composiciones, su transcripción –incluso utilizando herramientas de digitalización y OCR, que no son infalibles y requieren una revisión exhaustiva– y su posterior edición XML-TEI –solamente en lo que se refiere a las etiquetas estructurales– hubieran requerido un tiempo excesivo para los plazos que actualmente marcan los programas de doctorado.

En segundo lugar, el marcado semántico basado en modelos como los descritos en los párrafos anteriores se antoja imposible para un proyecto que no tenga detrás un equipo de trabajo. En el caso de desarrollar una metodología paralela a la de BIESES y SILEM, por ejemplo, la creación y validación de un esquema propia que refleje correctamente el contenido que se quiere resaltar –y las redes de etiquetas necesarias para traducir esa ontología a XML-TEI–, así como el ulterior marcado de los 549 textos recopilados para el estudio, sería una tarea dirigida más a un grupo de investigación que a un trabajo individual.

En tercer y último lugar, “faltan herramientas de producción ergonómicas que ahorren el

¹⁷ El lector interesado puede acceder a los buscadores y a su imprescindible guía de uso desde: <https://www.bieses.net/buscador.php> y <http://www.uco.es/investigacion/proyectos/silem/buscador/motor-busqueda.php>.

¹⁸ El propio equipo del proyecto Eros & Logos reconocía en su sesión de trabajo del congreso internacional *Venus a través del espejo: erotismo y creación en el mundo hispánico* –celebrado en la Universidad de Valladolid entre el 8 y el 11 de mayo de 2019– que la base de datos de textos, accesible desde: <http://www.erosylogos.com/>, no puede considerarse, por el momento, más que una transcripción, por lo que, a mi juicio, no puede servir como base para un análisis lingüístico de las características que se proponía en el estudio.

trabajo directo con el código informático” (Allés-Torrent, 2015, p. 19). En la práctica, esto suponía que, una vez terminada la hercúlea tarea de transcribir, editar y codificar más de quinientos textos, todavía sería necesario implementar un motor de búsqueda que permitiera recuperar la información lingüística requerida en cada momento. Ante la falta de un gestor de TEI que incluya una herramienta de búsqueda propia –como ocurre con gestores web, como WordPress o Joomla, o software, como Omeka o Atom–, la solución pasaría por desarrollar un buscador específico para el proyecto partiendo de cero, una labor totalmente inasumible para un filólogo y que requiere de personal informático especializado¹⁹.

2. METODOLOGÍA

La metodología elegida para el estudio fue una suerte de método híbrido entre la exégesis filológica tradicional –esto es, la lectura e interpretación de los textos– y lo que Allés-Torrent (2017) denomina “el método [...] de recolección de información en fichas” (p. 23), concretamente a partir de la creación de una base de datos relacional basada en el código abierto SQL²⁰.

La herramienta de análisis que se confeccionó para la preparación del estudio partía, pues, de las bases bibliográficas tradicionales, pero, lejos de quedarse en lo meramente descriptivo, se alimentó con información lingüística, poética, autorial y semántica, entre otras cuestiones.

A pesar de sus limitaciones, esta metodología mixta permitió solventar con éxito la mayor parte de los obstáculos descritos en los párrafos anteriores.

Primeramente, porque a partir de las distintas tablas y campos de descripción que se van interrelacionando, se pudo incluir en la base de datos toda la información “semántica” necesaria referida al *código erótico*. Así, como si de un etiquetado TEI se tratase, una búsqueda relacional permite, por ejemplo, recuperar la palabra en su forma estándar, “lema”, su aparición en un contexto concreto, “forma”, su significado en ese caso, “definición”, la composición en la que se encuentra, “composición”, y sus datos bibliográficos, “fuente” –entre otras muchas posibilidades y combinaciones–.

Estas pesquisas, al contrario de lo que ocurría con la codificación XML-TEI, no requieren de conocimientos informáticos especializados, pues el gestor de bases de datos PHPMyAdmin incluye un motor de búsqueda avanzado. Además, aunque para las búsquedas relacionales más complejas hay que recurrir directamente a expresiones de código SQL, estas pueden construirse de forma más o menos sencilla partiendo de búsquedas en la red o recurriendo puntualmente a algún especialista²¹.

¹⁹ Precisamente esto fue lo que ocurrió en el proyecto BIESES y SILEM, cuyo buscador fue desarrollado por un equipo informático externo.

²⁰ Esta base de datos se encuentra en el servidor de la Fundación Ramón Menéndez Pidal, a la que agradezco sinceramente su generosidad, accesible desde: <https://bit.ly/2lpDuQS>, con el usuario: *tesisPiquero* y la contraseña: *tesisPiquero*. Una vez dentro, se ha de seleccionar la base de datos *imaginaria_erotica_aurea*.

²¹ Debo agradecer en este punto la inestimable ayuda de Javier Nobel Antón, CEO de Peñalara Software, a la hora de escribir adecuadamente la sintaxis SQL, así como de conformar definitivamente la estructura relacional de la base de datos y la nomenclatura de las tablas y descriptores.

Frente a las completas ediciones TEI, este método ofrecía una limitación: para conocer el contexto léxico, el investigador ha de volver una y otra vez a la fuente bibliográfica en la que se encuentra editado el testimonio. El problema, sin embargo, era mínimo, pues, por un lado, entre los objetivos de la tesis no figuraba en ningún caso la edición de los textos, sino su interpretación; y por otro, cualquier investigador está más que habituado a la continua consulta directa de las fuentes a la hora de trabajar –incluso en los casos en los que se dispone de una edición digital–²².

En resumen, la recuperación de los datos textuales y contextuales ordenados a partir de tablas relacionadas resultó más que suficiente para la revisión e interpretación del lenguaje erótico que se pretendía abordar.

Dejando a un lado estas generalidades, conviene describir ahora cómo se estructuró internamente la base de datos a partir de dos tipos de tablas: las tablas entidad –ocho–, que recogen los datos principales de los que parte el análisis; y las tablas relacionadas –otras ocho–, que permiten representar las relaciones existentes con las primeras para su posterior confrontación y análisis.

Las ocho tablas entidad son²³:

- *Lema*. Término en su forma estándar –lover, caballo–.

Campos descriptores: id | lema | nota |

- *Forma*. Aparición concreta de la palabra en un testimonio –lloviendo–, incluyendo expresiones eróticas “complejas” del tipo *a caballo* o *acto cordial*. Una “forma”, pues, puede estar formada por uno o varios “lemas”, como ocurre por ejemplo en el caso extremo de *jugar de lomos y caderas*, donde cada término independiente no tiene sentido erótico por sí mismo, pero unidos forman la metáfora de ‘copular’²⁴.

Campos descriptores: id | forma | observaciones |

- *Definición*. Definición de una “forma” concreta en un contexto determinado.

Campos descriptores: id | forma_id | definicion²⁵ | observaciones |

- *Composición*. Información relativa al poema en el que se incluye la “forma”. Además de los datos referidos a la identificación del texto, se añaden detalles métricos y estróficos.

Campos descriptores: id | titulo | primer_verso | arte | metro | estrofa | poema | observaciones |

²² Además, hay que tener en cuenta que una base de datos relacional es un trabajo en permanente progreso, por lo que este pequeño escollo podría ser salvado próximamente a partir de la inclusión de un campo descriptor donde se incluya el texto completo, la cita concreta del poema o, de forma más avanzada, hipervinculando cada palabra a la digitalización correspondiente del poema.

²³ Como se verá, además de los descriptores específicos de cada tabla, todas ellas tienen un campo “id”, un número *currens* unívoco con el que se identifica cada dato para estructurar las relaciones, y otro “observaciones”, que permite al investigador añadir la información adicional que considere en cada caso.

²⁴ El total de “formas” indivisibles que poseen más de un “lema” suma 1045, aunque no todas ellas formaron parte del análisis final.

²⁵ Para evitar problemas de codificación SQL, se elimina la tilde en todos los campos descriptores que deberían aparecer tildados.

- *Autor.* Autor de la composición. El código VIAF (Virtual International Authority File) es un número de identificación unívoco que evita posibles redundancias y confusiones²⁶. Las fechas de nacimiento y muerte se consignan solo en caso de conocerse, si no, se recurre a un abanico de años o al siglo.

Campos descriptores: id | nombre | codigo_viaf | fecha_nacimiento | fecha_muerte |

- *Fuente.* Fuente bibliográfica en la que aparece la composición. Existen tres tipos: EM (edición manejada), F (fuente), que se refiere a los testimonios manuscritos y OE (otras ediciones), que consignan otras ediciones tanto antiguas como modernas de los textos²⁷. La cantidad de campos disponibles permite reflejar datos de manuscritos, monografías, artículos de libro, artículos de revista y cualquier otro formato. Los descriptores referidos a la cronología aluden, en general, a dos tipos de dato: por un lado, año, abanico de años o, en última instancia, siglo en el que se copia un manuscrito o se publica una edición antigua; por otro, año de nacimiento y muerte—exacto o aproximado— o siglo en el que desarrolló su actividad el autor editado²⁸. Las observaciones a esta cronología suelen consignar de dónde se recupera la información cronológica.

Campos descriptores: id | autor | titulo | en_titulo | volumen | editores | titulo_revista | numero | lugar | en_lugar | editorial | en_editorial | anyo | en_anyo | paginas | biblioteca | tipo | cronologia_inicio | cronologia_fin | cronologia_observaciones | observaciones |

- *Fuente_bibliografía.* Fuente bibliográfica secundaria, es decir, monografías, artículos y notas en las que apoyar las interpretaciones y definiciones de los “lemas”.

Campos descriptores: id | autor | titulo | en_titulo | volumen | editores | titulo_revista | numero | lugar | en_lugar | editorial | en_editorial | anyo²⁹ | en_anyo | paginas | observaciones |

- *Categoría.* Una serie de “palabras clave” o “vocabulario controlado” alusivos tanto a las prácticas sexuales como a los campos semánticos —e incluso a algunas cuestiones críticas— a partir de los cuales poder agrupar las imágenes eróticas. Esta tabla fue primordial para el estudio, puesto que el análisis léxico se estructuró a partir de órganos y prácticas sexuales y de campos semánticos.

²⁶ Cuando un autor no dispone de VIAF, el descriptor queda con un valor nulo.

²⁷ La intención de esta tabla fue consignar el mayor número de copias posibles de los poemas para futuros estudios o revisiones bibliográficas; no obstante, dado que la intención del trabajo era puramente hermenéutica y la edición no entraba en sus pretensiones, en ningún caso se consultaron ni cotejaron otras fuentes que no fuesen la edición manejada (EM) —salvo las variantes aparecían en las propias EM, como en el caso de *Poesía erótica del Siglo de Oro* (Alzieu, Robert y Lissorgue, 2000).

²⁸ Si se conoce la fecha exacta de producción de un autor, como en el caso de Góngora, se refleja este y no su fecha de nacimiento y muerte.

²⁹ En esta y la siguiente tabla, el campo “anyo” se debe escribir así para evitar errores de codificación de SQL con el carácter “ñ”.

Campos descriptores: id | categoría | observaciones |

En cuanto a las ocho tablas relacionadas, estas serían³⁰:

- *Forma_lema*. Relaciona cada “lema” con una aparición concreta, “forma”.

Campos descriptores: id | forma_id | lema_id | observaciones |³¹

- *Forma_composición*. Relaciona cada aparición, “forma”, con la “composición” concreta en la que aparece. Además, asocia esta aparición específica a una “definición”. Esta fue la tabla de trabajo principal durante el proyecto, puesto que refleja los datos principales del análisis semántico: palabra, ejemplo y significado.

Campos descriptores: id | forma_id | composicion_id | verso | fuente_manejada_id | verso_en_fuente³² | pagina_en_fuente | definicion_id | observaciones |

- *Composición_fuente*. Asocia cada “composición” a su “fuente”. En general, se añade información tanto de la Edición Manejada como de Fuentes y Otras Ediciones –cuando disponía de datos referidos a estas–.

Campos descriptores: id | composicion_id | fuente_id | ubicacion³³ | observaciones |

- *Autor_composición*. Relaciona cada “composición” con su “autor” –existiendo, obviamente, un autor “Anónimo”–. Hay dos tipos: F (fiable) y A (atribuido). En las “observaciones” de la tabla se refleja generalmente la fuente en la que se apoya esa atribución.

Campos descriptores: id | composicion_id | autor_id | tipo | observaciones |

- *Lema_categoría*. Relaciona cada “lema” con una o varias “categorías”. Habitualmente, cada “lema” tiene al menos dos: categoría sexual y categoría semántica.

Campos descriptores: id | lema_id | categoria_id | observaciones |

- *Lema_fuente_bibliografía*. Relaciona cada “lema” con uno o varios estudios en los que aparece citado.

Campos descriptores: id | lema_id | fuente_bibliografia_id | ubicacion |

³⁰ Es importante señalar que todas ellas poseen una cardinalidad N a M o relación “muchos a muchos”, lo cual implica que cada entrada de una tabla entidad puede estar relacionada con distintos datos de la entidad con la que se relaciona. Por ejemplo, una “forma” puede aparecer en varias “composiciones” y, a la vez, una “composición” está compuesta por varias “formas”.

³¹ En esta y las demás tablas relacionales se apunta siempre hacia el “id” del dato, ya sea lema, forma, composición, etc., puesto que recuperar la información a partir de las palabras, donde se puede confundir fácilmente un carácter, sería contraproducente para las búsquedas en la base de datos.

³² Algunas ediciones manejadas, como la del *Cancionero de obras de burlas provocantes a risa* (Bellón y Jauralde Pou, 1974), reinicia el número de verso en cada página, de manera que pareció lógico incluir en esta tabla tanto el número de verso de la composición –para que se pueda recuperar la referencia sea cual sea la edición usada– y el verso en la fuente manejada –para recuperar rápidamente la cita de la fuente manejada–.

³³ En esta y otras tablas, “ubicación” se refiere a la página o el folio dentro de cada testimonio.

observaciones |³⁴

- *Categoría_fuente_bibliografía*. Relaciona cada “categoría” con una o varias investigaciones que traten sobre ese asunto en concreto.

Campos descriptores: id | categoría_id | fuente_bibliografía_id | ubicacion | observaciones |

- *Autor_fuente_bibliografía*. Relaciona cada “autor” con uno o varios trabajos que traten sobre él.

Campos descriptores: id | autor_id | fuente_bibliografía_id | ubicacion | observaciones |³⁵

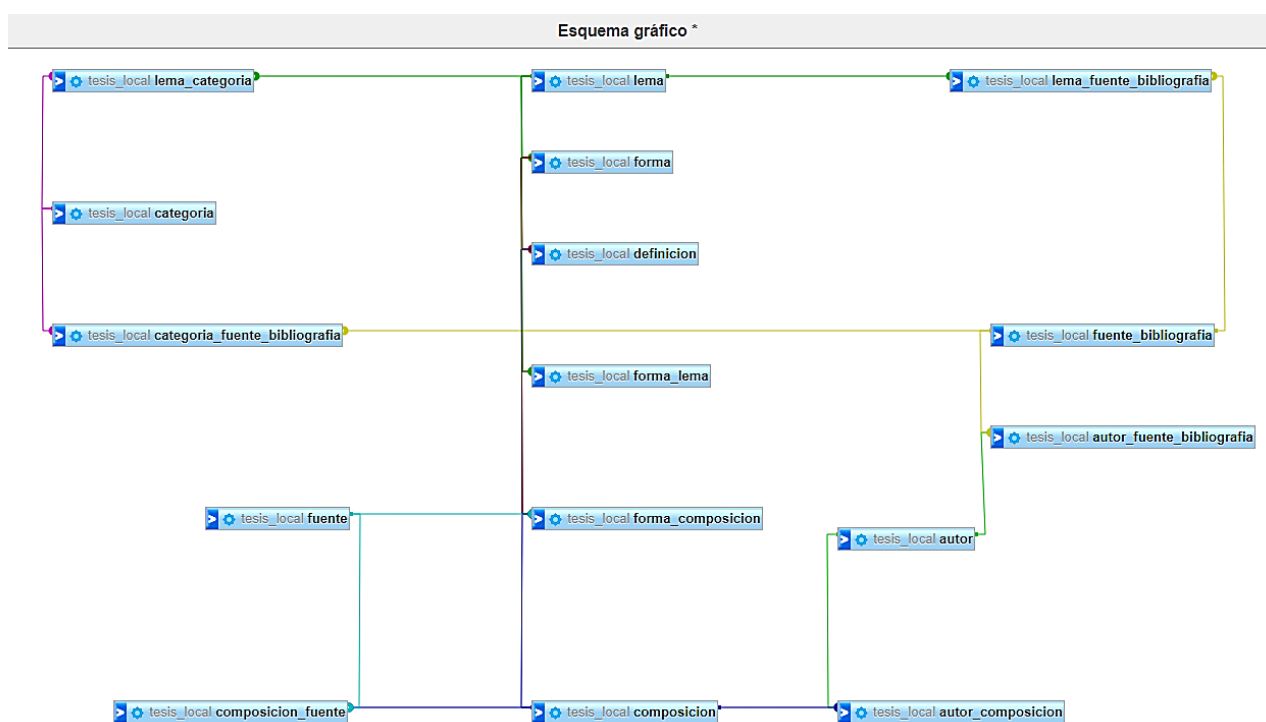


Figura 1. Esquema gráfico de las relaciones de la base de datos creado a partir de la herramienta Diseñador de PHPMyAdmin. Fuente: elaboración propia.

Una lectura atenta de la descripción de las tablas –y su posterior representación visual–, muestra que en la base de datos se consignaron tres clases de fichas distintas, aunque complementarias. En primer lugar, un bloque amplio referido propiamente al léxico erótico, es decir, el “lema”, la “forma”, la “definición”, la “composición”, el “autor” y la “fuente”. En segundo lugar, una serie de registros que sirvieron de herramienta para estructurar el posterior análisis, las “categorías”, referidas tanto a prácticas y usos sexuales como a campos semánticos o conceptos críticos. Por último, cuatro tablas complementarias dedicadas específicamente a las publicaciones científicas, que, como si de las clásicas fichas de lectura se tratase, permiten recuperar de un

³⁴ En este caso y los siguientes, el campo observaciones, de texto libre, posee una descripción más o menos amplia del documento bibliográfico concreto, incluyendo tanto resúmenes como citas literales. Esto facilitó notablemente la redacción del trabajo y permitirá recuperar información de las fuentes secundarias de una manera muy sencilla en estudios futuros.

³⁵ Como el punto de vista de este trabajo no es editorial, no se creó una tabla “composición_fuente_bibliografía” para recuperar los posibles trabajos referidos a un poema concreto. Sin embargo, queda abierta la

vistazo teorías e interpretaciones relevantes para el análisis.

A partir de todo lo anterior, a la hora de aplicar esta metodología a la redacción de un análisis comentado del vocabulario sexual se crearon cinco búsquedas combinadas en tres pasos, todas ellas enfocados a recuperar de la forma más clara y sintética posible la información requerida para cada apartado y subapartado del trabajo.

El primer paso en esta exploración de la base de datos se apoya en el filtrado y recuperación de los “lemas” por “categoría” y, sobre todo, por doble “categoría”. La estructuración de cada apartado se rigió fundamentalmente por los conceptos “campo sexual” y “campo semántico”, por tanto, esta doble indagación en los datos ofrece al usuario en pocos segundos una lista completa de palabras referidas a las dos “categorías” que le interesasen en cada momento. Así, buscando a la vez “genital masculino” y “bélico”, por ejemplo, el gestor de la base de datos recupera inmediatamente una lista de vocabulario fálico-bélico previamente marcado con la que comenzar a trabajar³⁶.

Una vez recuperada la lista de vocablos específica para cada subepígrafe –con la que se trabajaba después en el formato tradicional XLSX–, el siguiente paso, y sin duda el más importante, es el que permite recuperar en una única tabla todas las apariciones de cada uno de los “lemas” de la lista anterior junto con la información cuantitativa, semántica y bibliográfica disponible. Así, por ejemplo, al procurar *lanza*, en apenas unos segundos se puede conocer que tiene 29 apariciones totales, cuáles son sus distintas formas –*lança, lanza, lanza (quebrada)*–, qué significado tiene en cada ocasión –en este caso, ‘pene’–, en qué composiciones se utiliza el término (tanto el “título” como el “primer verso”), en qué versos concretos del poema, en qué fuente bibliográfica está el texto (autor, año, título), en qué verso de esa edición y en qué página.

Con esta búsqueda general, pues, el investigador tiene a su disposición de un solo vistazo todos los datos necesarios para construir un discurso analítico basado en las apariciones del vocablo en el corpus y para citar los ejemplos concretos sin necesidad de recurrir nuevamente a la edición física del poema³⁷.

Por último, y precisamente con el fin de recabar con la mayor brevedad y rigor todos los datos posibles para abordar con garantías la redacción de los numerosos epígrafes, con cada palabra –o campo de análisis– se llevaba a cabo una búsqueda de bibliografía secundaria basada en los dos descriptores anteriores: “categoría” y “lema”. En ambos casos la tabla de recuperación de datos permite conocer el autor y título del elemento bibliográfico concreto, las páginas de interés y la ficha de lectura o la definición.

Nuevamente, a través de estos dos sencillos pasos cualquiera podría conseguir toda la información crítica necesaria para desarrollar un análisis comparativo: por un lado, la búsqueda por “categorías” –genital masculino, sodomía, oficios, animales– ofrece un apoyo básico para el

posibilidad en un futuro.

³⁶ Evidentemente, en los campos semánticos limítrofes, en los que se puede generar algún desajuste, como por ejemplo en los “oficios” agrícolas y los “vegetales”, la búsqueda anterior se realizaba una vez con cada una de las categorías, confirmando de esta manera que el glosario estaba completo.

³⁷ Aunque en algunos casos dudosos es inevitable recurrir a una relectura de las composiciones.

planteamiento teórico de cada apartado del estudio; y por otro, la búsqueda por “lema” ofrece la (s) distinta(s) definiciones que la crítica ha dado a una voz en función de los diferentes contextos, metodologías y puntos de vista³⁸.

Una vez desglosada y organizada la ingente cantidad de información que se podía rastrear a partir de estas cinco búsquedas SQL, el investigador estaba ya en disposición de abordar con suficientes garantías el estudio léxico propuesto.

3. DISCUSIÓN

A pesar de su viabilidad técnica y teórica expuesta, es imprescindible señalar aquí algunos problemas estructurales y posibles mejoras para abordar en el futuro.

La idea primigenia del trabajo fue desarrollar un léxico de la poesía erótica de los Siglos de Oro al uso –similar al que ahora está disponible en la web del proyecto *Eros & Logos*–, por lo que la base de datos se confeccionó pensando que el dato nuclear a partir del cual se irían interrelacionando todos los demás sería “lema”. Tiempo después, sin embargo, para evitar coincidir con los objetivos del proyecto citado, la perspectiva metodológica del trabajo cambió, tomando forma la idea definitiva de abordar un análisis comentado del imaginario erótico áureo a partir de la división del vocabulario en campos semánticos y prestando especial atención a los datos cuantitativos para conocer su distribución y evolución.

En este nuevo escenario el “lema” continuó siendo la columna vertebral en la que se apoyaba todo el estudio, pues, más allá de reflejar la cantidad de apariciones distintas que puede tener tal o cual palabra, lo que se pretendía era mostrar la mayor o menor variabilidad que el vocabulario erótico posee en cada parcela de realidad. En este sentido, no resulta tan relevante que *lanza* tenga veinte o treinta apariciones en los distintos textos como que, dentro del campo semántico de la guerra, existan diez o quince instrumentos afines que cumplen la misma función metafórica y que demuestran la vitalidad y la originalidad de las disemias bélicas en el periodo.

Lo anterior, que tiene enormes ventajas analíticas, supuso también una serie de problemas que hubo que subsanar durante la redacción del ensayo. Dentro de ese nuevo objetivo marcado, la definición exacta de cada término quedaba relegada a un segundo plano, por lo que la construcción de la base de datos –como se deduce del esquema gráfico anterior– se reenforzó específicamente hacia la relación existente entre las “categorías” –sexuales o semánticas– y los “lemas”. ¿Cuál fue, entonces, el inconveniente? Que, lógicamente, al haberse construido una base de datos con relaciones de “muchos” a “muchos”, cada uno de estos “lemas” podía estar asociado

³⁸ Este volcado de lemas desde la bibliografía secundaria no incluyó los distintos glosarios eróticos disponibles (Criado de Val, 1960; Cela 1974; Vasvári, 1988; Reynal, 1988; Huerta Calvo, 1983, pp. 39-68; McGrady, 1984, pp. 105-108; Tello, 1992; Alzieu, Robert y Lissorgue, 2000, pp. 329-354; *Eros & Logos*, 2017-2021; Herrero Diéguez, Martínez Deyros y Sánchez Mateos, 2018, pp. 151-184, Garrote Bernal, 2020, pp. 245-272), pues, además de ser una labor inabarcable para la limitación temporal del trabajo, se entendía que cualquier lector puede recuperar fácilmente la información lematizada alfabéticamente en cada uno de ellos. Todo lo contrario ocurría con la bibliografía crítica, donde las referencias a tal o cual palabra se entrecruzan en los distintos párrafos sin un orden determinado.

a su vez a una o más “categorías”, lo que supone que el mismo término se pueda usar por ejemplo para referirse a la flacidez y a la impotencia o a la erección y a la flacidez si se menciona en contextos distintos —o incluso al pene y la vagina—.

En la práctica, esto se tradujo en que el total de la suma de los datos que aparecía en la tabla *lema_categoria* (5468) era obviamente mayor que el número total de “lemas” (2313). Teniendo en cuenta esta oscilación, en el análisis posterior no resultó posible reflejar unos datos porcentuales definitivos, sino aproximados. Así, desde el punto de vista cuantitativo, el estudio realizado debe considerarse una primera aproximación al léxico erótico a partir de una herramienta digital, pero nunca un esquema definitivo y cerrado del panorama³⁹.

Más allá de una futurible reestructuración, este pequeño escollo cuantitativo se solventó con la creación de una serie de gráficas anexas en las que sí se manejaron los datos exactos de “lemas” y “formas” totales citados en el análisis: 1133 lemas y 3549 apariciones. Si bien es cierto que cuantitativamente estos números quedan lejos del total de referencias que posee la base de datos, a mi juicio fue una muestra lo suficientemente amplia como para mostrar una visión muy cercana a la realidad del comportamiento del vocabulario sexual en la poesía erótica del periodo.

Siguiendo con la reflexión teórica sobre las posibles mejoras o aplicaciones de la herramienta en el futuro, los campos descriptores que no aparecen reflejados en ninguna de las búsquedas señaladas arriba obedecen a la intención de dejar abierta la posibilidad de abordar, al menos, tres tipos de análisis más: uno métrico, uno cronológico y uno autorial. Dadas las limitaciones temporales del proyecto y la extensión que finalmente ocupó el análisis léxico, estos apartados quedaron fuera de los objetivos del estudio, pero podrían recuperarse en una segunda fase.

En lo que respecta a la métrica, la idea parte de la inclusión en la tabla “composición” de notas detalladas acerca del arte, el metro, la clase de estrofa o el tipo de poema, de tal manera que, a través de una búsqueda SQL combinada con las “formas” —que nunca se terminó de desarrollar—, podría abordarse un análisis comparativo entre las imágenes de la poesía de tono popularizante y de la culta. Aunque es obvio que las diferencias entre una y otra vertiente no dependen exclusivamente del metro utilizado, partiendo de estos datos, y confrontándolos con otros referidos, por ejemplo, a los autores, se podrían comprobar cuestiones tan interesantes como si lo popular y lo culto comparten el mismo tipo de imaginaria o si una de las dos tradiciones influye claramente sobre la otra.

J. Ignacio Díez Fernández (2003), que ya se ocupó largamente del tema de los autores y lectores eróticos (pp. 42-62), señala que la división maniqueísta popular-culto no es efectiva cuando nos enfrentamos a esta clase de textos —ni a muchos otros del periodo—, pues “está fuera

³⁹ Como los medios digitales están permanentemente sujetos a revisiones, la recuperación exacta de estos datos podría ser abordada en un futuro próximo a partir de la modificación parcial de la estructura de tablas. Una de las posibles soluciones sería la creación de una tabla relacionada *categoria_definicion*, previa regularización de cada una de las definiciones, para recuperar unos datos cuantitativos mucho más exactos.

de discusión que existe un fondo común de saber erótico-literario, de muy variada procedencia, del que se nutren numerosos escritores” (pp. 57-58). A mi juicio, su afirmación es totalmente acertada, pero la posibilidad de demostrarla y ampliar la perspectiva crítica a partir de un análisis apoyado en la base de datos sería muy enriquecedora.

El segundo apartado pendiente, relativo a la cronología, es notablemente más incierto en cuanto a su metodología y los posibles resultados, puesto que tratar de asociar una fecha concreta a textos pocas veces impresos y continuamente maltratados en copias clandestinas se antoja una tarea casi imposible. A pesar de este grave obstáculo, la inclusión en la tabla “fuente” de información relativa al año de copia de los manuscritos –o al abanico de años⁴⁰, a la fecha de impresión de un cancionero o al periodo de vida de los autores –si no se posee ninguna de las anteriores– permitiría, cuando menos, tener un dato cronológico con el que proponer ciertas hipótesis. En este sentido, a partir de la confrontación de la “cronología” con los de “lemas” y “formas”, se podría arrojar nueva luz sobre las diferencias entre el vocabulario sexual de arraigo medieval –como el del *Cancionero de obras de burlas provocantes a risa* de 1519– y el posterior, así como concretar si la poética erótica tendió en el periodo áureo hacia la renovación o el conservadurismo⁴¹.

El último punto de los anunciados arriba, referido a los autores y atribuciones, es probablemente el más irrealizable de los tres, pero no por ello resulta menos atractivo. Idealmente, el análisis se apoyaría en una técnica similar a la estilometría, pero aplicada a partir de la información incluida en la base de datos. Así, se cotejarían los datos referidos a la “forma”, la “composición” y el “autor” para comprobar si una atribución puede ser más o menos fiable en función de la coincidencia de las imágenes eróticas. Indudablemente, para que la técnica fuera posible habría que incluir algún otro tipo de variable, ya que el vocabulario sexual es, en un alto porcentaje, tan universal que difícilmente se puede adscribir al *usus scribendi* de un escritor u otro. Sea como fuere, esta clase de búsquedas podrían mostrar resultados bastante interesantes y contribuirían a abrir nuevos caminos a la crítica de cara a la atribución de los poemas.

4. RESULTADOS

La metodología híbrida digital propuesta en esta nota de investigación permitió la recopilación y recuperación de datos léxicos de forma intuitiva, rápida y organizada para la redacción del trabajo de tesis doctoral *La imaginería en la poesía erótica de los Siglos de Oro* (Piquero, 2021), cuyos resultados pueden sintetizarse en tres puntos:

1. Análisis comentado de 1130 “lemas” y 3549 “formas” a partir de la lectura e

⁴⁰ Siempre y cuando se utilicen fuentes cercanas al periodo estudiado, que en este caso engloban los siglos XVI y XVII.

⁴¹ Una intensa lectura del corpus elegido para el trabajo, así como los datos recogidos en los distintos epígrafes, me lleva a sospechar que el constante uso del vocabulario explícito que aparece en composiciones tan provocadoras como la *Carajicomedia* fue poco a poco dejando paso al equívoco y las complejas dilogías en manos de poetas cultos como Baltasar del Alcázar, fray Melchor de la Serna o Luis de Góngora. Lo anterior, no obstante, no tendrá ninguna validez científica hasta que un estudio analítico de los datos lo corrobore.

interpretación de 549 poemas y más de 250 referencias de bibliografía secundaria. Todo ello supone un acercamiento al vocabulario erótico áureo desde una perspectiva amplia, sistemática y original que complementa a la perfección los estudios publicados hasta el momento.

2. Propuesta de un posible esquema estructural del lenguaje sexual a partir de las diez gráficas anexas al estudio.
3. Desarrollo de tres completos glosarios –alfabético, por campo sexual y por campo semántico– que sirven de herramienta fundamental para el investigador interesado en el léxico erótico.

En definitiva, la metodología de base de datos relacional SQL descrita en este artículo posee las garantías suficientes como para abordar cualquier estudio léxico de estas características, pues, una vez alimentada la herramienta con los datos oportunos y la estructura necesaria, se puede realizar una eficiente y rigurosa organización del material y una sencilla y minuciosa recuperación de la información.

Además, la creación de esta clase de metodologías digitales no solo ofrece resultados para un estudio concreto, sino que abren la posibilidad de continuar en el futuro con investigaciones desde otras vías y perspectivas⁴². En este sentido, aumentar el corpus, ampliar los límites cronológicos –hacia la imaginería erótica medieval o ilustrada– o desarrollar, a partir de la misma estructura, proyectos referidos a otra clase de género o de vocabulario, supone un mínimo esfuerzo para el investigador y presupone unos resultados muy prometedores.

REFERENCIAS BIBLIOGRÁFICAS

- Allés-Torrent, S. (2020). Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales? *Studia Aurea*, 14, 63-98. <https://doi.org/10.5565/rev/studiaaurea.395>
- Allés-Torrent, S. (2017). “Tiempo hay de acometer y tiempos de retirar”: literatura áurea y edición digital. *Studia Aurea*, 11, 13-30. <https://doi.org/10.5565/rev/studiaaurea.261>
- Allés-Torrent, S. (2015). Edición digital y algunas tecnologías aliadas. *Ínsula*, 822, 18-21.
- Alzieu, P., Robert J., & Lissorgues, Y. (Eds.). (2000). *Poesía erótica del Siglo de Oro*. Crítica.
- Bellón, J. A., & Jauralde Pou, P. (Eds.). (1974). *Cancionero de obras de burlas provocantes a risa*. Akal.
- Blasco, J. (2020). Antipetrarquismo. En J. Blasco, & C. Ruiz Urbón (Eds.), *Vocabulario del ingenio erótico en la poesía de los Siglos de Oro*. Eros & Logos (pp. 20-25). Peter Lang.
- Blasco, J. (2015). En el nombre de Venus: un “arte de amar” español del siglo XVI. En J. Blasco Pascual (Eds.), *Lasciva est nobis pagina... Erotismo y literatura española en los Siglos de Oro* (pp. 143-179). Editorial Academia del Hispanismo.
- Buzzetti, D. (2009). Digital Edition and Text Processing. En M. Deegan, & K. Sutherland (Eds.), *Text*

⁴² De hecho, hoy la cantidad de textos que conforman el corpus ha ascendido a 660 y se está empezando a trabajar en una base de datos con una estructura similar –aunque más refinada– para estudiar el erotismo en el romancero panhispánico.

- Editing, Print and the Digital World* (pp. 45-61). Ashgate.
- Cela, C. J. (1974). *Diccionario secreto*, 2 vols. Alianza Editorial.
- Criado de Val, M. (1960). Antífrasis y contaminaciones de sentido erótico en La lozana andaluza». *EnStudia Philologica. Homenaje a Dámaso Alonso* (pp. 431-457). Gredos.
- Débox, M. (1989). “Cogiendo rosas y lirios”. ¿Erotismo codificado?. En Covadonga López Alonso, Juana Martínez Gómez, José Paulino Ayuso, Marcos Roca y Carlos Sainz de la Maza (Coords.), *Eros literario. Actas del Coloquio celebrado en la Facultad de Filología de la Universidad Complutense en diciembre de 1988* (pp. 31-44). Universidad Complutense de Madrid.
- Díez Fernández, I. (2019). “Obscenidad de ideas y de palabras” en el Cancionero de obras de burlas provocantes a risa. En *Fiebre de luz río y de corceles. Poesía y erotismo áureo* (pp. 15-51). Medio Maravedí.
- Díez Fernández, I. (2003). *La poesía erótica de los Siglos de Oro*. Laberinto.
- Fradejas Rueda, J. M. (2009-2010). La codificación XML/TEI de textos medievales. *Memorabilia*, 12, 219-247. <http://uvadoc.uva.es/handle/10324/29211>
- Garrote Bernal, G. (2012). Practicantes del ingenio sexual (siglos XIII-XVII). *AnMal Electrónica*, 32, 235-175. http://www.anmal.uma.es/numero32/Ingenio_sexual.htm
- Garrote Bernal, G. (2020). *Con dos poéticas, teoría historicista de la literatura sexual española*. Editorial Agilice Digital.
- Herrero Diéguez, J., Martínez Deyros, M. & Sánchez Mateos, Z. (Eds.) (2018). “Aquel coger a oscuras a la dama”: mujeres en la poesía erótica del Siglo de Oro (antología), estudio preliminar de Patricia Marín Cepeda. Agilice Digital.
- Huerta Calvo, J. (1983). Cómico y femenino bureo (Del amor y las mujeres en los entremeses del Siglo de Oro). *Criticón*, 24, 5-68. https://cvc.cervantes.es/literatura/criticon/PDF/024/024_007.pdf
- Lacarra Lanz, E. (1996). Sobre los “dichos lascivos y rientes” en Celestina. En V. Roncero López, & A. Menéndez Collera (Coords.), “Nunca fue pena mayor”. *Estudios de literatura española en homenaje a Brian Dutton* (pp. 419-433). Ediciones de la Universidad de Castilla-La Mancha.
- McGann, J. (2004). Marking Texts of Many Dimensions. En S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *Companion to Digital Humanities* (pp. 198-217). Blackwell Publishing. https://companions.digitalhumanities.org/DH/?chapter=content/9781405103213_chapter_16.html
- McGrady, D. (1984). Notas sobre el enigma erótico, con especial referencia a los Cuarenta enigmas en lengua española. *Criticón*, 27, 71-108. https://cvc.cervantes.es/literatura/criticon/PDF/027/027_072.pdf
- Navarro-Colorado, B. (2015). A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects. En *Proceedings of the Fourth Workshop on Computational Linguistics for Literature* (pp. 105-113). Association for Computational Linguistics. <https://aclanthology.org/W15-0712>
- Piquero, Á. (2021). *La imaginería en la poesía erótica de los Siglos de Oro*. Tesis Doctoral.

- Universidad Complutense de Madrid. <https://eprints.ucm.es/id/eprint/69018/>
- Revenga, N. (2014). La edición crítica digital de los textos teatrales. La estrella de Sevilla como ejemplo. *Anuario de Lope de Vega. Texto, literatura, cultura*, XX, 99-121. <https://raco.cat/index.php/anuariolopedevega/article/view/286695>
- Reynal, V. (1988). *El lenguaje erótico medieval a través del Arcipreste de Hita*. Editorial Playor.
- Spence, P. (2014). La investigación humanística en la era digital: mundo académico y nuevos públicos. En A. Baraibar Echeverria (Coord.), *Humanidades digitales: una aproximación transdisciplinar* (pp. 117-131). Universidade da Coruña, SIELAE.
- Spence, P. (2015). Creación y difusión del conocimiento científico en la era digital. *Ínsula*, 822, 8-12.
- Tello, A. (1992). *Gran diccionario erótico de voces de España e Hispanoamérica*. Temas de Hoy.
- Vasvári, L. O. (1988). Vegetal-Genital Onomastics in the *Libro de Buen amor*. *Romance Philology*, 42(1), 1-29. <https://www.jstor.org/stable/44943967>
- Vélez, J. I., & Marmolejo-Ramos, F. (2016). Los secretos de *Cien Años de Soledad*: una aproximación estilométrica para la investigación en psicolingüística. *Revista colombiana de psicología*, 25.2, 265-288. <https://doi.org/10.15446/rcp.v25n2.50742>
- Whinnom, K. (1981). *La poesía amatoria cancioneril en la época de los Reyes Católicos*. University of Durham.

ANEXO. BÚSQUEDAS COMBINADAS ESPECÍFICAS PARA EL ANÁLISIS

1. Filtrar lema por categoría

```
SELECT l.lema, c.categoría
FROM lema_categoria AS lc
LEFT JOIN lema AS l ON lc.lema_id = l.id
LEFT JOIN categoría AS c ON lc.categoría_id = c.id
WHERE c.categoría = "oficios"
```

2. Filtrar lema que pertenece a dos categorías al mismo tiempo

```
SELECT l.lema, c.categoría
FROM lema AS l
LEFT JOIN lema_categoria AS lc ON l.id = lc.lema_id
LEFT JOIN categoría AS c ON lc.categoría_id = c.id
WHERE c.categoría = "oficios" AND l.id in (SELECT l.id from lema as l LEFT JOIN lema_categoria AS lc ON l.id = lc.lema_id LEFT JOIN categoría AS c ON lc.categoría_id = c.id WHERE c.categoría = "genital masculino")
```

3. Ficha completa del lema: lema, forma, definición, composición (título y primer verso), verso (y verso en fuente), página, fuente, autor de la fuente, año de la fuente, título de la fuente

```

SELECT FC.id, L.id as lemaid, L.lema, FC.forma_id as formald, F.forma, D.id as defld, D.definicion,
FC.verso, C.id as composicionId, C.titulo, C.primer_verso, FC.verso_en_fuente, FC.pagina_en_fuente,
FU.id as fuenteld, FU.autor, FU.anyo, FU.titulo
FROM forma_composicion AS FC
LEFT JOIN forma AS F ON FC.forma_id=F.id
LEFT JOIN lema AS L ON F.lema_id = L.id
LEFT JOIN definicion AS D ON FC.definicion_id = D.id
LEFT JOIN composicion AS C ON FC.composicion_id = C.id
LEFT JOIN fuente AS FU ON FC.fuente_manejada_id = FU.id
WHERE L.lema = 'lanza'

```

4. Filtrar por categoría y fuente bibliográfica

```

SELECT CAT.categoria, FB.id AS Id_FunteBib, FB.autor, FB.titulo, FB.anyo, FB.en_anyo,
CFB.ubicacion, CFB.observaciones
FROM categoria AS CAT, fuente_bibliografia AS FB, categoria_fuente_bibliografia AS CFB
WHERE CAT.categoria = 'semen' AND CAT.id = CFB.categoria_id AND FB.id =
CFB.fuente_bibliografia_id

```

5. Filtrar por lema y fuente bibliográfica

```

SELECT LEM.lema, FB.id AS Id_FunteBib, FB.autor, FB.titulo, FB.anyo, FB.en_anyo, LFB.ubicacion,
LFB.observaciones
FROM lema AS LEM, fuente_bibliografia AS FB, lema_fuente_bibliografia AS LFB
WHERE LEM.lema = 'jugar' AND LEM.id = LFB.lema_id AND FB.id = LFB.fuente_bibliografia_id

```