

## LINGÜÍSTICA DE CORPUS E HISTORIAS DE LAS LENGUAS IBERORROMÁNICAS: NUEVAS PROPUESTAS Y ÚLTIMOS DESARROLLOS

MIRIAM BOUZOUITA (*Humboldt-Universität zu Berlin*)

miriam.bouzouita@hu-berlin.de

ORCID-iD: <https://orcid.org/0000-0001-8196-620X>

ANTOINE PRIMERANO (*Universiteit Gent*)

antoine.primerano@ugent.be

ORCID-iD: <https://orcid.org/0000-0003-0437-5646>

### RESUMEN

Este artículo ofrece una introducción al número monográfico *Lingüística de corpus e historias de las lenguas iberorrománicas: nuevas propuestas y últimos desarrollos*, que contiene trabajos basados en contribuciones presentadas en el V Congreso Internacional de Corpus Diacrónicos en Lenguas Iberorrománicas (*CoDiLI5*), celebrado en la Universidad Libre de Bruselas (ULB) y coorganizado por la Universidad de Gante (UGent). Esta introducción consta de dos partes. En primer lugar se exponen los principales enfoques e intereses de los encuentros *CoDiLI* y, específicamente, de la quinta edición (Bruselas 2019), así como las relaciones que existen entre los trabajos aquí recogidos y estos enfoques. Luego, se le dedica a cada uno de estos trabajos un resumen de su temática, su(s) pregunta(s) de investigación y sus principales resultados.

**PALABRAS CLAVE:** lingüística de corpus, historias de las lenguas iberorrománicas, *CoDiLI*

### CORPUS LINGUISTICS AND THE HISTORIES OF THE IBERO-ROMANCE LANGUAGES: NEW PROPOSALS AND RECENT DEVELOPMENTS

#### ABSTRACT

This article offers an introduction to the thematic issue *Corpus Linguistics and the Histories of the Ibero-Romance Languages: New Proposals and Recent Developments*, which contains papers based on talks given at the V Congreso Internacional de Corpus Diacrónicos en Lenguas Iberorrománicas (*CoDiLI5*), held at the Free University of Brussels (ULB) and co-organised by Ghent University (UGent). This introduction consists of two parts. Firstly, we summarise the main foci and interests of the *CoDiLI* conferences and, more specifically, of its fifth edition (Brussels 2019), as well as the links between the papers included in this volume and these foci. Secondly, for each paper a short summary of its topic is given, alongside the treated research question(s) and main results.

**KEY WORDS:** corpus linguistics, histories of the Ibero-Romance languages, *CoDiLI*

#### 1. INTRODUCCIÓN

En las últimas dos décadas se ha presenciado el tremendo incremento en la creación de herramientas lingüísticas para el análisis histórico de las lenguas y variedades iberorrománicas, como los extensos bancos de datos (p.ej. CORDE), así como en los estudios llevados a cabo con corpus automatizados, creando por tanto la necesidad de establecer un foro para poder compartir resultados científicos y discutir problemas metodológicos asociados al uso y a la construcción de tales instrumentos computacionales. La serie internacional de congresos de *Corpus Diacrónicos en Lenguas Iberorrománicas (CoDiLI)* nació precisamente de esta necesidad. La gran mayoría de las contribuciones incluidas en este número monográfico se presentaron en el V Congreso Internacional de esta serie (*CoDiLI5*), celebrado en la Universidad Libre de Bruselas y coorganizado por la

Universidad de Gante, los días 12-14 de diciembre de 2019. Como en los encuentros anteriores (Mallorca 2007, Barcelona 2011, Zúrich 2014, Alcalá de Henares 2016), el propósito central de esta edición era doble: pretendía, por un lado, dar a conocer los últimos desarrollos y avances en el campo de la construcción de corpus diacrónicos para las lenguas iberorrománicas y, por otro, crear un espacio de debate sobre la utilización de tales corpus. Además, la edición celebrada en 2019 tuvo un enfoque particular, ya que también se centraba, por una parte, en las lenguas iberorrománicas minoritarias (p.ej. el leonés, el aragonés, el judeoespañol, etc.) y el habla de grupos minoritarios (p.ej. el lenguaje de esclavos, moriscos, etc.) y, por otra, en el contacto entre lenguas y variedades diatópicas en las diferentes etapas de la historia de los iberorromances. Como se verá en más detalle en el siguiente apartado, ambas vertientes temáticas específicas están ampliamente representadas en este número. En breve, lo ejemplifican el trabajo de Egidio Fernández sobre las huellas lingüísticas asturleoneras y aragonesas en documentación del siglo XVII, el de Taberero Sala sobre la elaboración de un corpus del habla navarra de los siglos XVI-XIX, las contribuciones de Arnold y Quintana, que presentan diferentes herramientas tecnológicas para el estudio (de la historia) del judeoespañol, así como el artículo de Vivancos Mulero sobre el habla veratense, en cuyo léxico se observan los efectos del contacto del andaluz con el dialecto murciano por las repoblaciones históricas, o el de Marcet Rodríguez, quien describe un nuevo corpus centrado en la variedad de Ávila, en la cual confluyen rasgos norteños y meridionales del español peninsular. De la misma manera, el estudio de Pineda indaga en la historia del catalán la existencia de un fenómeno morfosintáctico tradicionalmente vetado debido a su supuesto estatus de interferencia española, es decir, producto del contacto del catalán con el español. Las contribuciones de Calderón Campos y Vaamonde Dos Santos, y de Codita asimismo atañen al tema central del *CoDiLi5*, puesto que aquella da a conocer al público la creación de un nuevo corpus diacrónico centrado en el español peninsular moderno, y esta resalta un área descuidada en la investigación hispánica con la ausencia de corpus para el español colonial panameño.

El encuentro 2019 contó con cinco comunicaciones plenarias, impartidas por Cristina Egidio Fernández (Universidad de León), Mónica Castillo Lluch (Université de Lausanne), Lola Pons Rodríguez (Universidad de Sevilla), y Patricia Rochwert-Zuili (Université d'Artois) y Hélène Thieulin-Pardo (Sorbonne Université), así como una mesa redonda sobre las complicaciones metodológicas relacionadas con la construcción de corpus históricos centrados en el habla de minorías, en la cual participaron Mar Garachana (Universitat de Barcelona), Beatriz Arias Álvarez (Universidad Nacional Autónoma de México), Mercedes Abad Merino (Universidad de Murcia) y Charlotte Galves (Universidade de Campinas). Aparte de ello, se presentaron 45 diferentes comunicaciones y participaron presencialmente más de 50 científicos.

## 2. EL PANORAMA DE LAS CONTRIBUCIONES

Abrimos este número monográfico con la contribución de una de las plenaristas del *CoDiLi5*, **María Cristina Egidio Fernández** (Universidad de León), que tiene como objetivo determinar el grado de vitalidad de los rasgos dialectales asturleonéses y aragoneses en documentos notariales del s. XVII procedentes respectivamente de Asturias y León, y de Aragón. El estudio empírico llevado a cabo para ello se basa en datos procedentes del

*Corpus Léxico de Inventarios (CorLexIn)* y demuestra, entre otros, que la documentación notarial se presta perfectamente como fuente para el estudio histórico de características lingüísticas. Como es sabido, el asturleonés y el aragonés, lenguas hoy arrinconadas a áreas geográficas de superficie reducida en el extremo norte peninsular, gozan en la Edad Media de una amplitud geográfica, social y textual mucho mayor que la que conocen en la actualidad. Para ilustrarlo, se presentan primero las más destacadas características medievales de los dos romances, tanto en el plano fonético (p.ej. los diptongos decrecientes /ei/ y /ou/, el resultado /j/ de los grupos -LY-, -K'L-, -G'L- y la evolución PL-, KL-, FL- > /ʎ/-/tʃ/-/ʝ/ para el asturleonés; la apócope extrema de -e y -o, el resultado /it/ de los grupos -KT- y -ULT- y el resultado /ʎ/ de los grupos -LY-, -K'L-, -G'L- para el aragonés) como en el gramatical (p.ej. los resultados *yo(s)*, *lla(s)*, *ye/lle(s)* de los pronombres de objeto ILLUM/-OS, ILLAM/-AS, ILLI(S) para el asturleonés; el posesivo de tercera persona plural *lur(es)* para el aragonés). A partir del siglo XIII, ambos romances se hallan sometidos a un proceso de castellanización que acaba eliminándolos casi por completo de la escritura en el siglo XV, bien que en el registro oral logren sobrevivir (a duras penas) hasta la actualidad.

Aparte de examinar e ilustrar ampliamente los rasgos medievales del asturleonés y aragonés, se proporcionan también muestras orales contemporáneas de las dos lenguas, caracterizadas hoy por algún castellanismo fruto del proceso de asimilación lingüística anteriormente mencionado. Visto el marcado retroceso sufrido por estas dos lenguas entre la Edad Media y la actualidad, el artículo pretende establecer el grado de vitalidad de estos rasgos diatópicamente marcados en una etapa intermedia, a saber, el siglo XVII, a partir de relaciones de bienes contenidas en el *CorLexIn*. La autora demuestra claramente que estos documentos del siglo XVII, pese a ser de base claramente castellana, siguen exhibiendo rasgos dialectales que se han mantenido vivos hasta esa época, como los diptongos decrecientes /ei/ y /ou/ del asturleonés, el resultado /it/ < -KT-, -ULT- del aragonés, y las soluciones /j/ (asturleonés) y /ʎ/ (aragonés) que provienen de los grupos -LY-, -K'L-, -G'L-.

Pasamos al centro de la Península Ibérica en el segundo artículo de este número monográfico, el cual da a conocer la creación de una nueva herramienta para el estudio histórico del español, a saber, el *Corpus de Documentos de Ávila de la Hispanic Society of America (CODAHS)*. Esta contribución pretende realizar una descripción de los documentos incluidos en este corpus, todos redactados en la provincia de Ávila en los siglos XV y XVI, y hoy en día conservados en el Hispanic Museum and Library, una institución estadounidense que promueve el arte y la cultura de los países de habla hispana y portuguesa en Estados Unidos y contiene más de 250.000 documentos redactados desde el siglo XI hasta la actualidad. Los documentos que conforman el CODAHS se editan bajo la dirección de **Vicente J. Marcet Rodríguez** (Universidad de Salamanca-IEMYRhd), también el autor de esta contribución. La región de Ávila resulta de gran interés lingüístico por ser una zona relativamente poco estudiada hasta los años 90 del siglo pasado, a pesar de situarse en el confluente del español norteño y del español meridional y de presentar unos rasgos característicos que ponen de manifiesto la variación que encierra el español de Castilla y León, como por ejemplo la aspiración y la pérdida de la /-s/ implosiva o la conservación de una aspirada o un sonido velar /h-/ procedente de la F- inicial latina. De momento CODAHS se compone de seis documentos reales del siglo XV, así como de un cartulario relacionado con los bienes de la catedral de Ávila, otro cartulario y otros documentos jurídicos relacionados con el monasterio de San Jerónimo de Guisando. El

artículo empieza por describir el corpus en cuestión, indicando, entre otras cosas, la datación, la ubicación y el tema de cada documento considerado. A continuación se realiza un exhaustivo análisis paleográfico que revela la presencia a lo largo del corpus de tres letras (gótica cursiva, cortesana y humanística), las transiciones históricas entre una y otra y, en el caso de esta última, la adopción de determinados patrones escriturarios. Sigue un estudio gráfico-fonológico que deja vislumbrar ciertos cambios diacrónicos y/o características propias del habla de la zona, como (i) la aspiración y la pérdida de la F- latina, (ii) la aspiración y la pérdida de /-s/ implosiva, y (iii) la coalescencia de los pares de sibilantes apicoalveolares /s/-/z/, dentoalveolares /ts/-/dz/ y prepalatales /ʃ/-/ʒ/.

El tercer artículo, de la mano de **Cristina Tabernero Sala** (Universidad de Navarra), nos lleva a las variedades norteñas centrales y presenta una propuesta para la elaboración de un corpus lingüístico para la Navarra de los siglos XVI-XIX, un área y un periodo hasta la fecha descuidados en los estudios diacrónicos del español, aunque la variedad románica medieval de esta región esté bastante bien estudiada. Este enfoque, no obstante, no carece de interés, pues, por un lado, esta región participa en el *continuum* del español norteño y en la variación que encierra y, por tanto, necesita de estudios que identifiquen sus principales características, y, por otro lado, se caracteriza por un prolongado contacto histórico con la lengua vasca, el cual dejó sus huellas en la variedad diatópica de la zona a través de mecanismos de transferencia. El corpus propuesto se centra en una tipología documental en particular: las declaraciones de litigantes y testigos en procesos judiciales, específicamente los procesos por heridas y malos tratos del Archivo Diocesano de Pamplona y los pleitos por injurias del Archivo General de Navarra, si bien también incluye unas cartas privadas y otros documentos jurídicos. Estas tipologías documentales presentan la ventaja de acercarse al polo de la inmediatez comunicativa, polo que permite detectar los rasgos diatópicos diferenciales, y cuya utilidad se ha comprobado en estudios dialectológicos recientes (cf. los trabajos de Egido Fernández y Tabernero Sala en este volumen) y en la creación de corpus centrados en ello (cf. Calderón Campos y Vaamonde Dos Santos en este volumen). Además de una descripción y justificación del corpus en cuestión, la autora lleva a cabo un estudio lingüístico pormenorizado de los documentos, en el cual distingue los siguientes elementos característicamente navarros presentes en los documentos, la mayoría de ellos atribuibles a una impronta vascuense: (i) las grafías <yl(l)> e <yn(n)> para representar los fonemas /l/ y /ɲ/ respectivamente, (ii) la indistinción entre los fonemas /s/ y /θ/, y (iii) la presencia intensa de un léxico romance típico de la región (p.ej. *vanovero*, *fiemo*, *muir*, *costiero*), así como de ítems léxicos vascos (p.ej. *baçarre*, *laya*) insertados en discursos pronunciados/escritos en español, a manera de cambio de código (*code-switching*).

Desplazamos el punto de enfoque desde el iberorromance central, con la propuesta de la creación de un corpus navarro, hacia las hablas orientales de la Península Ibérica con la contribución de **Anna Pineda** (Sorbonne Université), en la que se examina en la historia del catalán la existencia del marcaje diferencial de objeto (MDO), un fenómeno bien conocido en varias lenguas, entre ellas el español. Primero, la autora indaga las descripciones normativas del fenómeno, desde Fabra (1918 [1933]) hasta la última gramática (GIEC 2016), para determinar cómo las perspectivas prescriptivistas sobre esta característica morfosintáctica han evolucionado. Concluye que, a pesar de que en la

tradición prescriptiva se ha ido aumentando el número de contextos permitidos para la aparición del MDO, existe aún una divergencia importante entre lo que la gramática normativa establece y lo que ocurre en el catalán diario, ya que gran parte de los casos del habla espontánea sigue censurada, como por ejemplo el MDO ante un sintagma nominal con referente humano o con un nombre propio no dislocado y/o no ambiguo (p.ej. *He vist a la teva cosina; He vist a la Maria*). A través de ejemplos extraídos de varios tipos de fuentes escritas y orales, se demuestra la vitalidad contemporánea del MDO catalán, tanto en registros formales como informales, que incluso llega a afectar a las categorías inferiores de las escalas de definitud y animacidad, como los indefinidos y los inanimados (p.ej. *Els determinants acompanyen als noms*). Bien que tradicionalmente se considere el MDO en catalán como una interferencia española, se evidencia con base en un estudio diacrónico extenso con datos del *CICA* que se trata de una evolución interna, es decir, de un rasgo propio del catalán, puesto que también se manifiesta en los siglos XIII-XV, etapas anteriores a la presión sociolingüística del español, que surge a finales del siglo XV y a principios del siglo XVI. Ahora bien, la autora demuestra que, aunque el español no parece ser la fuente de este cambio en catalán, sí parece haber ejercido una influencia cuantitativa en el desarrollo del MDO catalán y parece haber reforzado un proceso de gramaticalización iniciado antes del contacto.

Con el siguiente artículo pasamos de las hablas iberorrománicas orientales a las sudorientales, ya que este trabajo trata la confección de un pequeño corpus especializado para el estudio diacrónico y diatópico de la frontera oriental de las hablas andaluzas en los siglos XVI y XVII. En especial, la autora de esta contribución, **Esther Vivancos Mulero** (Universidad de Murcia), se centra en el examen del habla de la Tierra de Vera, una zona de la denominada «Andalucía murciana», área adscrita geográficamente al territorio andaluz, pero cuyas características dialectales permiten su inclusión en las variedades murcianas. Para poder determinar la extensión geográfica y cronológica de la isoglosa del murciano por la Andalucía oriental, se elabora un corpus que comprende cuarenta y dos protocolos notariales para el siglo XVI, y treinta y nueve para el siglo XVII. De estos se selecciona de manera equitativa para ambos siglos una muestra representativa de doscientos documentos, entre los cuales encontramos (i) cartas de dote y arras, (ii) inventarios de bienes, (iii) testamentos, (iv) almonedas, y (v) particiones de bienes.

Tras detallar la creación de este corpus andaluz oriental y la ventaja de los protocolos notariales como fuentes idóneas para el estudio histórico del vocabulario español, por ejemplo los inventarios de bienes que recogen parte del léxico de la vida cotidiana (cf. el corpus *ODE* descrito en Calderón Campos y Vaamonde Dos Santos en este volumen, y el *CorLexIn* en Egido Fernández en este volumen para la utilidad lingüística de este tipo documental), se estudian cuatro voces dialectales de este corpus, a saber, *arambre*, *cetra*, *tenaja*, *orón*, y sus variantes. Estos análisis léxicos pretenden corroborar la hipótesis según la cual las hablas andaluzas orientales se configuran como una zona de influencia del dialecto murciano por la repoblación histórica. Efectivamente, los resultados del estudio de estos cuatro vocablos manifiestan la presencia de orientalismos léxicos en el habla veratense que la diferencian de las demás hablas andaluzas: en concreto, *arambre* y *tenaja* parecen ser de origen aragonés, *cetra* de descendencia catalana y *orón* es un murcianismo. La examinación dialectal revela que todos fueron introducidos a través de las variedades

murcianas. Por consiguiente, se concluye que existe desde la primera repoblación del siglo XVI una clara influencia diatópica murciana en esta variedad de la Andalucía oriental.

En la siguiente contribución, también centrada parcialmente en el español andaluz, **Miguel Calderón Campos y Gael Vaamonde Dos Santos** (Universidad de Granada) presentan un nuevo corpus especializado para el estudio de la Edad Moderna, *Oralia diacrónica del español (ODE)*, de interés no solo para filólogos y lingüistas, sino también para paleógrafos e historiadores. Este corpus recoge documentación manuscrita inédita de tres tipos textuales que se acercan, de una manera u otra, al habla oral y cotidiana, a saber, (i) los inventarios de bienes (para la utilidad lingüística de esta fuente documental, véanse también los trabajos de Egido Fernández y Vivancos Mulero en este volumen), (ii) las declaraciones de testigos de procesos judiciales (cf. el corpus navarro propuesto por Tabernero Sala en este volumen), y (iii) las certificaciones periciales de cirujanos y sangradores asociadas a estos procesos judiciales. Tras motivar la decisión de incluir estas tres fuentes documentales en *ODE*, los autores pasan a explicar las diferencias y similitudes entre este corpus y su antecesor, el *Corpus diacrónico del español del reino de Granada, 1492-1833 (CORDEREGRA)*.

Bien que *ODE* sea la continuación del *CORDEREGRA*, difieren en varios aspectos elementales de índole geográfica, cronológica y metodológica. Primero, se amplía el alcance espacial de los documentos comprendidos. Mientras que el *CORDEREGRA* ofrecía documentos que provienen del reino de Granada, jurisdicción territorial de la Corona de Castilla desde la caída del reino nazarí hasta 1833, que comprende las actuales provincias de Granada, Málaga y Almería, el *ODE*, además de esto, contiene dos subcorpus de control, uno de la zona centro-norte peninsular y otro de la región andaluza occidental (Cádiz, Huelva y Sevilla), que permiten llevar a cabo estudios contrastivos con los datos granadinos. Segundo, tomando en consideración la extensión geográfica y, por ello, la pérdida del enfoque en el reino de Granada, que cesa de existir tras la reforma territorial de 1833, se extiende el arco temporal al siglo XIX entero, incluyendo de esta manera la Edad Moderna y los inicios de la Edad Contemporánea. Por último, se suma al rigor filológico que ya caracterizaba a *CORDEREGRA* la implementación de nuevos avances tecnológicos en *ODE*. Concretamente, para la edición digital de los textos, se transcriben primero los documentos en formato XML de acuerdo con la codificación del consorcio *Text Encoding Initiative (TEI)* y se procesan luego en la plataforma *TEITOK* (Janssen 2014) para la obtención de anotaciones morfosintácticas, la lematización y la normalización de las variantes.

Gran parte de este artículo se centra en este tercer aspecto, en que se dan a conocer los diferentes problemas metodológicos que presentaba el *CORDEREGRA* y se logran solucionar en *ODE*. Para ilustrar estos aspectos, mediante ejemplos concretos Calderón Campos y Vaamonde Dos Santos subrayan la necesidad de emplear en la creación de corpus, entre otros, un normalizador para poder enfrentarse a la variabilidad ortográfica de los documentos antiguos (para la misma conclusión respecto del judeoespañol, véase Quintana en este volumen). La implementación en *ODE* de los diferentes avances tecnológicos ofrecidos por el uso del lenguaje XML-TEI y la plataforma *TEITOK* resultan en un corpus más provechoso, ya que es anotado y permite, además de la recuperación de varios tipos de datos lingüísticos y metatextuales, tres visualizaciones de los documentos editados: (i) la edición facsímil, (ii) la paleográfica y (iii) la normalizada.

Sigue el artículo de **Rafael Arnold** (Universität Rostock), que presenta el proyecto *Diccionario del Español Medieval electrónico (DEMel)*, financiado por la Fundación Alemana para la Investigación Científica (DFG). Tiene como objetivo la digitalización del fichero del *Diccionario del Español Medieval (DEM)*, un proyecto de archivo léxico del español medieval iniciado en los años 1970 por Bodo Müller e interrumpido en 2007, tras el corte de ayudas financieras. Además de proporcionar una herramienta de gran calidad en los estudios del español medieval, ya dio frutos en su breve vida puesto que logra precisar las dataciones de gran cantidad de voces. Como explica el autor, el resultado del proyecto *DEMel*, concebido como un corpus complementario de los ya existentes, no solo consiste en poner a disposición de los usuarios en formato digital las aproximadamente 1.700 fichas que servían para la elaboración del *DEM*, sino también en crear un corpus de más de 30.000 lemas, con sus variantes gráficas y contextos de uso, que se puede emplear para los estudios lexicológicos diacrónicos. En este aspecto se destaca el *DEMel* por constituir una base de datos lematizada y etiquetada que se centra exclusivamente en el español medieval –mientras que otros corpus (p.ej. *CORDE*, *CNDHE*, etc.) abarcan un periodo mucho más extenso– y ofrece una presentación cronológica de las atestaciones de los vocablos recogidos. Tras describir los objetivos del *DEMel*, así como las características de su base material, a saber, el *DEM*, el autor nos informa sobre el avance del proyecto y las diversas etapas de su realización: (i) la digitalización y la descripción de las fichas, (ii) la lematización, y (iii) la integración de una interfaz de búsqueda. Finalmente, se demuestra cómo el uso del *DEMel* puede ser provechoso para el estudio histórico del judeoespañol, identificando no solo el origen, sino también la evolución de las voces españolas medievales que han sobrevivido en el judeoespañol actual, después de la expulsión de España (p.ej. *apiadar*, *afermosear*, *ambezar*).

Siguiendo con el tema del judeoespañol, **Aldina Quintana** (The Hebrew University of Jerusalem) expone en su contribución la creación, el desarrollo y el estado actual del *Corpus Diacrónico Anotado del Judeoespañol (CoDiAJe)*. Este proyecto pretende crear una base de datos digital amplia del judeoespañol que responda a las muchas dificultades inherentes a la documentación en esa lengua iberorrománica: (i) el empleo de diversos alfabetos –latino, cirílico, griego, pero sobre todo hebreo–, (ii) el alto grado de variación diatópica y diafásica, (iii) las adaptaciones del alfabeto hebreo a la fonología judeoespañola, y (iv) la presencia aguda de préstamos de otras lenguas. A la hora de redacción este corpus contiene setenta y cuatro documentos de los siglos XIV a XXI escritos en judeoespañol, que pueden visualizarse en cinco formas gráficas diferentes: (i) la grafía original, (ii) la transcripción latina, (iii) la forma expandida, (iv) el judeoespañol normalizado y (v) el español actual. La elaboración del corpus se hizo con la ayuda del programa *TEITOK* (Janssen 2014, cf. Calderón Campos y Vaamonde Dos Santos en este volumen), que ha permitido solucionar problemas derivados de la presencia del hebreo, no solo en las expresiones y citas hebreas incluidas en los documentos a modo de cambio de código (*code-switching*), sino también en las formas híbridas que caracterizan el léxico y la morfología del judeoespañol (p.ej. sustantivos hebreos en genitivo, palabras hebreas con morfemas gramaticales hebreos, auxiliares judeoespañoles con participios hebreos). Además de ofrecer información textual de este tipo, el corpus también provee información lingüística a través de (i) la lematización –llevada a cabo a partir del modelo de la variedad judeoespañola de Estambul, la que cuenta con el mayor número de hablantes–, (ii) el

etiquetado gramatical –a partir de las etiquetas de *EAGLES* para el español–, (iii) información semántico-conceptual, así como (iv) metadatos. Esto significa que el corpus permite buscar una gran cantidad de lemas y variantes asociadas a través de un motor de búsqueda *CQP*, sin perder de vista el alfabeto originario de los documentos.

El número monográfico se cierra con el artículo de **Viorica Codita** (Université de Neuchâtel), en el que se revisa críticamente la existente documentación archivística transcrita para el estudio del español colonial panameño, una de las variedades menos examinadas en el estudio histórico-diacrónico de la lengua española. Como se explica en esta contribución, esta situación se debe a la ausencia de un corpus delimitado compuesto de diferentes tipos de fuentes transcritas con criterios rigurosos (cf. red de proyectos *CHARTA*). Por ello, la mayoría de los estudios históricos sobre el español de Panamá solo ofrece breves descripciones, de corto alcance y poco profundas. La autora demuestra que las carencias de la documentación panameña disponible son de índole variada. Aunque algunas obras ofrecen transcripciones de documentos procedentes de los siglos *xvi* y *xvii*, su objetivo etnohistórico original las hace poco adecuadas para análisis lingüísticos. Otros trabajos, aparte de basarse en un número muy reducido de documentos, solo cubren la documentación del siglo *xvi* y *xviii*, desatendiendo por completo el siglo *xvii*, una época esencial para la historia del español latinoamericano, en la cual ya se observa una tendencia a adoptar construcciones diferentes según las diversas áreas hispanohablantes (Egido Fernández 2016). Asimismo, no existe documentación transcrita adecuadamente para el siglo *xix*. Por consiguiente, la autora se plantea la elaboración de un corpus transcrito con criterios rigurosos y fiables lingüísticamente que tome en cuenta tanto el criterio geográfico (a pesar de la escasa densidad poblacional de algunos territorios), como el étnico-social, dados los diferentes estratos de la sociedad colonial panameña, compuesta por españoles (y otros europeos), criollos, indígenas y africanos.

## REFERENCIAS BIBLIOGRÁFICAS

- ARNOLD, Rafael D. (en este volumen): «La digitalización del fichero del *Diccionario del Español Medieval (DEM)*: una nueva fuente para la historia del español y del judeoespañol», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum Digital*, 9, pp. 191-207.
- CALDERÓN CAMPOS, Miguel y Gael VAAMONDE DOS SANTOS (en este volumen): «*Oralia Diacrónica del Español*: un nuevo corpus de la Edad Moderna», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum digital*, 9, pp. 167-189.
- CHARTA = RED CHARTA: *Corpus Hispánico y Americano en la Red: Textos Antiguos*. <https://www.redcharta.es/> [Consulta: 15/12/2020].
- CICA = TORRUELLA, Joan (dir.): *Corpus Informatitzat del Català Antic*, elaborado con Manuel Pérez Saldanya y Josep Martines. [www.cica.cat](http://www.cica.cat) [Consulta: 15/12/2020].
- CNDHE = REAL ACADEMIA ESPAÑOLA (2013): *Corpus del Nuevo Diccionario Histórico del Español*. <http://web.frl.es/CNDHE/view/inicioExterno.view> [Consulta: 15/12/2020].
- CODAHSA = MARCET RODRÍGUEZ, Vicente J. (dir.): *Corpus de Documentos de Ávila de la Hispanic Society of America*.
- CODITA, Viorica (en este volumen): «La documentación de archivo y la historia del español panameño», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e*

- historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum digital*, 9, pp. 237-251.
- CoDIAJE = QUINTANA, Aldina (dir.): *Corpus Diacrónico Anotado del Judeoespañol*. <http://corptedig-glif.upf.edu/teitok/codiaje/> [Consulta: 15/12/2020].
- CORDE = REAL ACADEMIA ESPAÑOLA: *Corpus Diacrónico del Español*. <http://corpus.rae.es/cordenet.html> [Consulta: 15/12/2020].
- CORDEREGRA = *Corpus diacrónico del español del Reino de Granada (1492-1833)* <http://corpora.ugr.es/ode/> [Consulta: 15/12/2020].
- CorLexIn = MORALA RODRÍGUEZ, José Ramón (dir.): *Corpus Léxico de Inventarios (CorLexIn)*, <http://web.frl.es/CORLEXIN.html> [Consulta: 15/12/2020].
- DEM = MÜLLER, Bodo (1987-2005): *Diccionario del Español Medieval*, vol. 1, fascículos 1-10, vol. 2, fascículos 11-20, vol. 3, fascículos 21-26. Heidelberg: Winter.
- DEMel = ARNOLD, Rafael D. (dir.): *Diccionario del Español Medieval electrónico (DEMel)*. [disponible a partir de finales de 2020].
- EGIDO FERNÁNDEZ, María Cristina (2016): «América vs. España: contrastes gramaticales y léxicos en documentación del siglo XVII», en Marta Fernández Alcaide, Elena Leal Abad y Álvaro S. Octavio de Toledo y Huerta (eds.), *En la estela del Quijote: Cambio lingüístico, normas y tradiciones discursivas en el siglo XVII*. Fráncfort: Peter Lang, pp. 189-213.
- EGIDO FERNÁNDEZ, María Cristina (en este volumen): «Variación diatópica en documentos notariales del s. XVII: asturleonés y aragonés», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum digital*, 9, pp. 15-59.
- FABRA, Pompeu (1918 [1933]): *Gramàtica catalana*. Barcelona: Institut d'Estudis Catalans.
- GIEC = INSTITUT D'ESTUDIS CATALANS (2016): *Gramàtica de la llengua catalana*. Barcelona: Institut d'Estudis Catalans.
- JANSSEN, Maarten (2014): *TEITOK – a Tokenized TEI environment*. <http://www.teitok.org/> [Consulta: 15/12/2020].
- MARCET RODRÍGUEZ, Vicente J. (en este volumen): «El corpus de documentos de Ávila del Hispanic Museum and Library (siglos XV y XVI). Descripción paleográfica y gráfico-fonológico», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum digital*, 9, pp. 61-85.
- ODE = CALDERÓN CAMPOS, Miguel y María Teresa GARCÍA GODOY (2010-2020): *Oralia Diacrónica del Español*. <http://corpora.ugr.es/ode/> [Consulta: 15/12/2020].
- PINEDA, Anna (en este volumen): «Surgimiento y expansión del marcado diferencial de objeto en la diacronía del catalán», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum Digital*, 9, pp. 117-144.
- QUINTANA, Aldina (en este volumen): «CoDiAJE – The Annotated Diachronic Corpus of Judeo-Spanish. Description of a Multi-Alphabetic Corpus and its Textual and Linguistic Annotations», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum Digital*, 9, pp. 209-236.
- TABERNERO SALA, Cristina (en este volumen): «Contribución al estudio del español norteño a partir de un corpus de declaraciones en procesos judiciales (siglos XVI-XIX)», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum Digital*, 9, pp. 87-115.
- VIVANCOS MULERO, Esther (en este volumen): «La elaboración de un corpus de pequeño dominio para la caracterización dialectal del oriente andaluz», en Miriam Bouzouita y Antoine Primerano (eds.), *Lingüística de corpus e historias de las lenguas iberorrománicas: Nuevas propuestas y últimos desarrollos*, *Scriptum Digital*, 9, pp. 145-166.