



# PHILOLOGIA HISPANA ESTUDIOS LINGÜÍSTICOS



# PHILOLOGIA HISPALENSIS

AÑO 2024  
VOL. XXXVIII/1

ESTUDIOS LINGÜÍSTICOS



FACULTAD DE FILOLOGÍA  
UNIVERSIDAD DE SEVILLA

**EVALUACIÓN DE ORIGINALES:** Los originales se someten a una evaluación ciega, un proceso anónimo de revisión por pares, siendo enviados a evaluadores externos y también examinados por los miembros del Consejo de Redacción y/o los especialistas del Consejo Asesor de la Revista.

**PERIODICIDAD:** Anual en formato tradicional y en formato electrónico.

**PUBLICACIÓN EN INTERNET:** <<https://editorial.us.es/es/revistas/philologia-hispalensis>>, <<https://revistascientificas.us.es/index.php/PH>>.

**BASES DE DATOS:** *Philología Hispalensis* se encuentra indexada en CARHUS Plus+2018, CIRC (grupo B), DIALNET, DOAJ, Dulcinea, Index Islamicus, Latindex 2.0 (100% de los criterios cumplidos), MIAR (ICDS 2022 = 10), MLA, REDIB, SCOPUS, ERIHPLUS y ANVUR (Clase A). Asimismo, cuenta con el sello de calidad de la FECYT (8<sup>a</sup> edición, 2023) en los campos de conocimiento Lingüística y Literatura dentro de la modalidad Humanidades.

**ENVÍO DE ORIGINALES Y SUSCRIPCIONES:** Las colaboraciones deben enviarse a través de <<https://revistascientificas.us.es/index.php/PH>>.

**DIRECCIÓN DE CONTACTO:** Secretariado de la Revista Philología Hispalensis, Facultad de Filología, Universidad de Sevilla, C/ Palos de la Frontera, s/n, 41004 Sevilla; o bien al correo electrónico <[philhisp@us.es](mailto:philhisp@us.es)>.

**INTERCAMBIOS O CANJES (BIBLIOTECAS UNIVERSITARIAS):** Solicítense a Editorial Universidad de Sevilla o al Secretariado de la revista <[philhisp@us.es](mailto:philhisp@us.es)>.

© Editorial Universidad de Sevilla

Financiación: Revista financiada por la Universidad de Sevilla dentro de las ayudas del VII PPIT-US y del Decanato de la Facultad de Filología.

PORTADA: [referencias.maquetacion@gmail.com](mailto:referencias.maquetacion@gmail.com)

DEPÓSITO LEGAL: SE-354-1986

ISSN: 1132 - 0265 / eISSN 2253-8321

Maquetación: [referencias.maquetacion@gmail.com](mailto:referencias.maquetacion@gmail.com)

IMPRIME: Podiprint

DISTRIBUYE: Editorial Universidad de Sevilla, Porvenir, 27, 41013 Sevilla

Licence Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)



FECYT 641/2023  
Fechado el 01/07/2023 a las 18:00:00 (8<sup>a</sup> convocatoria)  
Vigencia hasta el 31 de julio de 2024



**EQUIPO EDITORIAL**

Directora:	Yolanda Congosto Martín, Universidad de Sevilla, España
Secretaria:	Leyre Martín Aizpuru, Universidad de Sevilla, España
Editoras:	Salomé Lora Bravo, Universidad de Sevilla, España
	Natalia Silva López, Fundación Pública de Estudios Universitarios “Francisco Maldonado” de Osuna, España
Coordinadora de Reseñas:	M. Amparo Soler Bonafont, Universidad Complutense de Madrid, España
Consejo de Redacción:	Gema Areta Marigó, Universidad de Sevilla, España Elisabetta Carpitelli, Université Stendhal - Grenoble Alpes, France Antonio Luis Chaves Reino, Universidad de Sevilla, España Marianna Chodorowska-Pilch, University of Southern California, USA Yves Citton, Université Paris 8 Vincennes-Saint Denis, France Ninfa Criado Martínez, Universidad de Sevilla, España María Dolores Gordón Peral, Universidad de Sevilla, España Isabel María Íñigo Mora, Universidad de Sevilla, España Manuel Maldonado Alemán, Universidad de Sevilla, España Daniela Marcheschi, Università degli Studi di Perugia, Italia Pedro Martín Butragueño, Colegio de México, México Miguel Ángel Quesada Pacheco, Universitetet i Bergen, Norge Angelica Valentinetti, Universidad de Sevilla, España Alf Monjour, Universität Duisburg-Essen, Deutschland María José Osuna Cabezas, Universidad de Sevilla, España Fátima Roldán Castro, Universidad de Sevilla, España Antonio Romano, Università degli Studi di Torino, Italia Juan Pedro Sánchez Méndez, Université de Neuchâtel, Suisse María Luisa Siguán Bochmer, Universitat de Barcelona, España José Solís de los Santos, Universidad de Sevilla, España Modesta Suárez, Université de Toulouse-Le Mirail, France María Ángeles Toda Iglesia, Universidad de Sevilla, España José Agustín Vidal Domínguez, Universidad de Sevilla, España María Jesús Viguera Molins, Universidad Complutense de Madrid, España Adamantía Zerva, Universidad de Sevilla, España

**COMITÉ CIENTÍFICO**

Juan Francisco Alcina Rovira, Universitat Rovira i Virgili, España
Gerd Antos, Martin-Luther-Universität Halle-Wittenberg, Deutschland
Gianluigi Beccaria, Università degli Studi di Torino, Italia
Isabel Carrera Suárez, Universidad de Oviedo, España
Carmen Herrero, Manchester Metropolitan University, England
Anna Housková, Univerzita Karlova, Česká Republika
Dieter Kremer, Universität Trier, Deutschland
Xavier Luffin, Vrije Universiteit Brussel, Belgique
Roberto Nicolai, Sapienza - Università di Roma, Italia
Marie-Linda Ortega, Université Sorbonne Nouvelle - Paris 3, France
Deborah C. Payne, American University, USA
Carmen Silva-Corvalán, University of Southern California, USA
Alicia Yllera Fernández, UNED, España

**CONSEJO ASESOR****ESTUDIOS ÁRABES E ISLÁMICOS**

Eva Lapiedra Gutiérrez, Universidad de Alicante, España  
Pablo Beneito Arias, Universidad de Murcia, España  
Carmelo Pérez Beltrán, Universidad de Granada, España

**FILOLOGÍA ALEMANA**

Georg Pichler, Universidad de Alcalá, España  
Marta Fernández-Villanueva Jané, Universitat de Barcelona, España  
María José Domínguez, Universidade de Santiago de Compostela, España

**FILOLOGÍA CLÁSICA - LATÍN**

Jesús Luque Moreno, Universidad de Granada, España  
José Luis Moralejo Álvarez, Universidad de Alcalá de Henares, España  
Eustaquio Sánchez Salor, Universidad de Extremadura, España

**FILOLOGÍA CLÁSICA - GRIEGO**

Didier Marcotte, Université Sorbonne Paris, France  
Maurizio Sonnino, Sapienza-Università di Roma, Italia  
Stefan Schorn, Université Catholique de Louvain, Belgique

**FILOLOGÍA FRANCESA**

Dolores Bermúdez Medina, Universidad de Cádiz, España  
Monserrat Serrano Mañes, Universidad de Granada, España  
María Luisa Donaire Fernández, Universidad de Oviedo, España

**FILOLOGÍA ITALIANA**

Giovanni Albertocchi, Universitat de Girona, España  
Cesáreo Calvo Riquel, Universitat de València - IULMA, España  
Margarita Borreguero Zuloaga, Universidad Complutense de Madrid, España

 **LENGUA ESPAÑOLA**

Emilio Montero Cartelle, Universidad de Santiago de Compostela, España  
Antonio Salvador Plans, Universidad de Extremadura, España  
Antonio Briz Gómez, Universitat de València, España

 **LENGUA INGLESA**

Emilia Alonso Sameño, Ohio University, USA  
Carmen Gregori Signes, Universitat de València, España  
Nuria Yanez-Bouza, Universidade de Vigo, España

 **LINGÜÍSTICA**

Ángel López García, Universitat de València, España  
Eugenio Martínez Celdrán, Universitat de Barcelona, España  
Juan Carlos Moreno Cabrera, Universidad Autónoma de Madrid, España

 **LITERATURA ESPAÑOLA**

Pedro M. Cátedra, Universidad de Salamanca, España  
Flavia Gherardi, Università degli Studio di Napoli Federico II, Italia  
Leonardo Romero Tobar, Universidad de Zaragoza, España

 **LITERATURA HISPANOAMERICANA**

Teodosio Fernandez, Universidad Autónoma de Madrid, España  
Noé Jitrik, Universidad de Buenos Aires, Argentina  
Edwin Williamson, Oxford University, Inglaterra

 **LITERATURA INGLESA**

Luis Alberto Lázaro Lafuente, Universidad Alcalá de Henares, España  
Ricardo Mairal Usón, UNED, España  
Carme Manuel Cuenca, Universitat de València, España

 **TEORÍA DE LA LITERATURA**

José Domínguez Caparrós, UNED, España  
Antonio Garrido Domínguez, Universidad Complutense de Madrid, España  
Isabel Paraíso Almansa, Universidad de Valladolid, España

**REVISORES DEL VOLUMEN 38, NÚMERO 1 (2024). ESTUDIOS LINGÜÍSTICOS**

Han actuado como revisores anónimos para uno o más artículos de este número, tanto los aceptados como los rechazados, los siguientes investigadores:

Alejo Fernández, Francisco (Universidad de Extremadura)  
Almeida Cabrejas, Belén (Universidad de Alcalá)  
Ariolfo, Rosana (Università degli Studi di Trieste, Italia)  
Cabello Pino, Manuel (Universidad de Huelva)  
Castillo Carballo, Auxiliadora (Universidad de Sevilla)  
Castillo Lluch, Mónica (Université de Lausanne, Suiza)  
Contreras Izquierdo, Narciso (Universidad de Jaén)  
Enrique-Arias, Andrés (Universitat de les Illes Balears)  
Felú-Arquiola, Elena (Universidad de Jaén)  
García Aranda, María Ángeles (Universidad Complutense de Madrid)  
Garrido Martín, Blanca (Universidad de Sevilla)  
Harjus, Jannis (Universität Innsbruck, Austria)  
Lorenzo Bergillos, Francisco José (Universidad Pablo de Olavide)  
Mariotini, Laura (Sapienza Università di Roma, Italia)  
Montoro del Arco, Esteban Tomás (Universidad de Granada)  
Morant i Marco, Ricard (Universitat de València)  
Moustaoui Srhir, Adil (Universidad Complutense de Madrid)  
Muñoz-Basols, Javier (Universidad de Sevilla)  
Paredes García, Florentino (Universidad de Alcalá)  
Perea Siller, Francisco Javier (Universidad de Córdoba)  
Quilis Merín, Mercedes (Universitat de València)  
Robles, Sara (Universidad de Málaga)  
Romera Manzanares, Ana María (Universidad Complutense de Madrid)  
Salvador Plans, Antonio (Universidad de Extremadura)  
Samper-Hernández, Marta (Universidad de Las Palmas de Gran Canaria)  
Santamaría Pérez, Isabel (Universidad de Alicante)  
Santana Marrero, Juana (Universidad de Sevilla)  
Torres Martínez, Marta (Universidad de Jaén)  
Ueda, Hiroto (Universidad de Tokio, Japón)



## ÍNDICE

<b>Sección Monográfica. Estudios diacrónicos y sincrónicos del paisaje lingüístico hispánico / <i>Diachronic and Synchronic Studies on the Hispanic Linguistic Landscape</i>.....</b>	<b>13</b>
Presentación / <i>Presentation</i> .....	15-20
Mercedes de la Torre García (Universidad Pablo de Olavide de Sevilla)	
La narración polifónica del estallido social chileno en el paisaje lingüístico de Santiago / <i>The Polyphonic Narration of the Chilean Social Uprising in the Linguistic Landscape of Santiago</i> .....	21-48
Maria Vittoria Calvi (Università degli Studi di Milano)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.i01.01">https://dx.doi.org/10.12795/PH.2024.v38.i01.01</a>	
Paisaje lingüístico y crematonomia: la estratificación social de los nombres comerciales / <i>Linguistic Landscape and Crematonymy: the Social Stratification</i> .....	49-68
Carmen Fernández Juncal (Universidad de Salamanca)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.i01.02">https://dx.doi.org/10.12795/PH.2024.v38.i01.02</a>	
Paisaje semiótico virtual de las enfermedades autoinmunes: el caso del lupus / <i>Virtual and Semiotics Landscape of Autoimmune Diseases: the Case of Lupus</i> .....	69-86
M.ª Victoria Galoso Camacho (Universidad de Huelva)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.i01.03">https://dx.doi.org/10.12795/PH.2024.v38.i01.03</a>	
El turismo en Huelva a través del paisaje lingüístico urbano / <i>Tourism in Huelva throughout the Urban Linguistic Landscape</i> .....	87-123
María Heredia Mantis (Universidad de Granada)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.i01.04">https://dx.doi.org/10.12795/PH.2024.v38.i01.04</a>	
El paisaje lingüístico religioso en Andalucía occidental: tipos de manifestaciones / <i>The Religious Linguistic Landscape of Western Andalusia: Types of Manifestations</i> .....	125-151
Francisco Molina-Díaz (Universidad Pablo de Olavide)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.i01.05">https://dx.doi.org/10.12795/PH.2024.v38.i01.05</a>	

Corpus PLANEO: metodología y resultados de un corpus de paisaje lingüístico andaluz / <i>PLANEO Corpus: Methodology and Results of a Corpus of Andalusian Linguistic Landscape</i> .....	153-166
Lola Pons Rodríguez (Universidad de Sevilla)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.06">https://dx.doi.org/10.12795/PH.2024.v38.io1.06</a>	
Una aproximación al estudio del paisaje lingüístico político en el País Vasco: análisis comparativo entre Vitoria-Gasteiz y Zizurkil / <i>An Approach to the Study of the Political Linguistic Landscape in the Basque Country: Comparative Analysis between Vitoria-Gasteiz and Zizurkil</i> .....	167-183
Urtzi Reguero Ugarte (Euskal Herriko Unibertsitatea, UPV/EHU)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.07">https://dx.doi.org/10.12795/PH.2024.v38.io1.07</a>	
Relaciones posibles entre temporalidad y paisaje lingüístico en español / <i>Possible Relationships between Temporality and Linguistic Landscape in Spanish</i> .....	185-212
Daniel M. Sáez Rivera (Universidad de Granada)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.08">https://dx.doi.org/10.12795/PH.2024.v38.io1.08</a>	
<b>Sección Varia</b>	
Pronombres personales y género gramatical en español / <i>Personal Pronouns and Grammatical Gender in Spanish</i> .....	215-238
Pedro Pablo Devís Márquez (Universidad de Cádiz)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.09">https://dx.doi.org/10.12795/PH.2024.v38.io1.09</a>	
Juan Gorgues y Lerma. Estudio historiográfico de la fraseología en <i>Lluvia de refranes</i> (1879) / <i>Juan Gorgues y Lerma. A Historiographic Study of Phraseology in Lluvia de refranes</i> (1879).....	239-260
Carmen Martín Cuadrado (Universidad Complutense de Madrid)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.10">https://dx.doi.org/10.12795/PH.2024.v38.io1.10</a>	
Percepción y reconocimiento de la diversidad lingüística del español por parte de jóvenes universitarios vietnamitas / <i>Perception and Recognition of Linguistic Diversity of Spanish Among Vietnamese University Students</i> ....	261-278
Ngọc Kim Ngân Võ (Universidad de Ciencias Sociales y Humanidades, Universidad Nacional de la Ciudad Ho Chi Minh)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.11">https://dx.doi.org/10.12795/PH.2024.v38.io1.11</a>	
La ironía en la fraseodidáctica: el caso de las locuciones verbales en estudiantes rusos de ELE / <i>Irony in the Phraseodidactics: the Case of Verbal Locutions in SFL Russian Students</i> .....	279-300
Pablo Ramírez Rodríguez (RUDN University)	
<a href="https://dx.doi.org/10.12795/PH.2024.v38.io1.12">https://dx.doi.org/10.12795/PH.2024.v38.io1.12</a>	

Anotación y explotación de variantes gráficas de base fonética en el corpus *Oralia Diacrónica del Español / Annotation and Exploitation of Phonetically-Based Spelling Variation in the Corpus Oralia Diacrónica del Español* ..... 301-323  
 Miguel Calderón Campos y Gael Vaamonde (Universidad de Granada)  
<https://dx.doi.org/10.12795/PH.2024.v38.i01.13>

**Reseñas de libros**

Soledad Chávez Fajardo: *Diccionarios del fin del mundo*. Santiago: Fondo de Cultura Económica Chile, 2022, 227 pp. ISBN: 978-956-289-248-3 ..... 327-331  
 Tania Avilés (School for International Training, SIT)

María Isabel Rodríguez Ponce: *Mitologías de la lingüística. Reflexiones sobre comunicación no sexista y libertad discursiva*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, 2022, 230 pp. ISBN: 978-84-9192-293-3 (Iberoamericana) / 978-3-96869-336-1 (Vervuert) ..... 333-336  
 Moisés Llopis i Alarcón (Universidad de Chile)

Raquel González Rodríguez y Cristina Sánchez López: *La negación en contraste*. Madrid: Arco/Libros, 2022, 181 pp. ISBN: 978-84-7133-878-5 ..... 337-341  
 Elena Monge Hermida (Universidad Complutense de Madrid)

David Malinowski, Hiram H. Maxim y Sébastien Dubreil: *Language Teaching in the Linguistic Landscape. Mobilizing Pedagogy in Public Space*. Cham, Suiza: Springer, 2020, 370 pp. ISBN: 2215-1656 ..... 343-345  
 Raquel Navas (University of Warwick)

Victoriano Gaviño Rodríguez y Miguel Silvestre Llamas (Eds.): *De eruditos, maestros, polemistas y otras figuras en la prensa del XIX. Estudios sobre la lengua y su enseñanza*. Madrid: Editorial Síntesis, 2023, 378 pp.  
 ISBN 978-84-1357-285-7 ..... 347-351  
 Andrea María Requena Millán (Universidad de Córdoba)

María Victoria Galloso Camacho, Manuel Cabello Pino y María Heredia Mantis (Eds.): *Funciones y aplicación didáctica del paisaje lingüístico andaluz*. Iberoamericana Vervuert, 2023, 261 pp. ISBN: 978-84-9192-313-8 ..... 353-359  
 María de las Mercedes Soto Melgar (Universidad de Jaén)





ESTUDIOS LINGÜÍSTICOS

**ANOTACIÓN Y EXPLOTACIÓN DE VARIANTES GRÁFICAS DE BASE FONÉTICA  
EN EL CORPUS *ORALIA DIACRÓNICA DEL ESPAÑOL***

**ANNOTATION AND EXPLOITATION OF PHONETICALLY-BASED SPELLING VARIATION  
IN THE CORPUS *ORALIA DIACRÓNICA DEL ESPAÑOL***

MIGUEL CALDERÓN CAMPOS

*Universidad de Granada*

calderon@ugr.es

ORCID: 0000-0002-0656-3643

GAEL VAAMONDE

*Universidad de Granada*

gaelvaamonde@ugr.es

ORCID: 0000-0001-8360-2805

Recibido: 31-07-2023

Aceptado: 22-02-2024

**RESUMEN**

Los corpus históricos de referencia disponibles para el español son de gran utilidad para investigar aspectos de carácter léxico y morfosintáctico, pero resultan inadecuados para extraer información de tipo fonético. Conscientes de esta necesidad, se están creando en los últimos años corpus especializados que prestan especial cuidado en el rigor gráfico de las transcripciones. El objetivo de este trabajo es doble. Por un lado, proponemos algunas líneas de mejora en la recolección de datos históricos que sean fiables para realizar estudios fonéticos. Por otro lado, presentamos el corpus especializado *Oralia Diacrónica del Español*, prestando especial atención a los aspectos que convierten a ODE en una herramienta idónea para abordar estudios cuantitativos sobre fonética histórica. Usando los datos de ODE, analizamos un estudio de caso, el de las grafías indicadoras de seseo y ceceo, con el objeto de demostrar la utilidad de este corpus para la investigación en fonética histórica del español clásico y moderno.

Palabras clave: corpus histórico, anotación de corpus, fonética histórica, seseo, ceceo.

**ABSTRACT**

Historical reference corpora available for Spanish are very useful for investigating lexical and morphosyntactic aspects, but they are inadequate for extracting phonetic information.

Aware of this need, in recent years specialized corpora have been developed that pay particular attention to the graphic accuracy of the transcriptions. The aim of this work is two-fold. On the one hand, we propose some lines of improvement in the collection of historical data that are reliable for carrying out phonetics approaches. On the other hand, we present a specialized corpus *Oralia Diacrónica del Español*, paying special attention to those aspects that make ODE a suitable tool to handle quantitative studies on historical phonetics. Based on the data in this corpus, we analyze a case study, that of the spellings indicating seseo and ceceo, in order to prove the usefulness of this corpus for research in the historical phonetics of classical and modern Spanish.

Keywords: historical corpus, corpus annotation, historical phonetics, seseo, ceceo.

## 1. INTRODUCCIÓN

Desde la aparición de los primeros corpus informatizados, la investigación lingüística basada en corpus se ha visto indiscutiblemente favorecida por los avances tecnológicos en el mundo de la computación, materializados en la creciente capacidad para almacenar y procesar datos lingüísticos de una forma que era poco menos que inimaginable hace pocas décadas. En este sentido, la lingüística de corpus se ha desarrollado a un ritmo realmente vertiginoso: en apenas sesenta años hemos pasado de tener corpus crudos de un millón de palabras accesibles a unos pocos usuarios a contar con corpus anotados compuestos por miles de millones de formas y puestos a disposición de cualquier persona con conexión a internet (Rojo, 2016: 285-286).

Esta revolución tecnológica, que ha derivado en la construcción de corpus textuales cada vez más provechosos para el usuario lingüista, también es trasladable a los corpus históricos. El CdEhist (Davies, 2001), con cien millones de palabras, o el más reciente CDH (Real Academia Española, 2013), que supera los 350 millones de palabras al momento de redactar estas líneas, constituyen una buena muestra de ello en el ámbito hispánico, al ser dos corpus diacrónicos anotados, de gran tamaño y de libre acceso en la red. Ambos recursos cuentan con interfaces de búsqueda de fácil manejo y han demostrado ser extraordinariamente útiles para obtener datos cuantitativos sobre fenómenos de carácter léxico y gramatical en diferentes períodos del español (Davies, 2009; Rojo, 2021). En términos de utilidad, sin embargo, los dos corpus citados cuentan con una limitación importante, que es aplicable también a otros corpus históricos de referencia, como el CORDE (Real Academia Española, s. f.): ninguno de ellos resulta adecuado para abordar estudios de fonética histórica.

Por fortuna, en el ámbito hispánico contamos hoy con varios corpus históricos de carácter especializado que facilitan el análisis de la variación gráfico-fonética. Cabe destacar en este sentido la valiosa aportación llevada a cabo por la red internacional CHARTA (Almeida Cabrejas, s. f.), coordinada por Pedro Sánchez-Prieto Borja, y en la que colaboran actualmente una treintena de equipos de investigación. La triple presentación de documentos que caracteriza a todos los corpus incluidos en

CHARTA —facsimilar, paleográfica y crítica—, así como el esfuerzo que ha promovido esta red por unificar criterios en lo tocante a la tarea de transcripción de fuentes documentales, han supuesto un punto de inflexión en la investigación histórica de fenómenos sobre fonética, grafémica y escriptología. Junto a CHARTA, y ciñéndonos a corpus que incluyen documentación del español clásico y moderno, contamos con los datos que aportan corpus especializados como CODEA+2022 (Grupo de investigación de textos para la historia del español, s. f.), CorLexIn (Morala Rodríguez, 2018a), Corpus Mallorca (s. f.), P. S. Post Scriptum (2014) o Oralía Diacrónica del Español (Calderón Campos y García-Godoy, 2019-). En torno a todos ellos, en los últimos diez años se ha renovado el interés por los estudios gráfico-fonéticos basados en documentación histórica, como muestran, entre otros, los trabajos siguientes: Almeida Cabrejas (2014); Vázquez Balonga y Sánchez-Prieto (2015); Almeida Cabrejas (2017); Ueda (2019); Morala y Perdiguero (2019); Sánchez-Prieto Borja y Vázquez Balonga (2019); Sánchez-Prieto Borja (2020); Ortiz Cruz (2020); Sánchez-Prieto Borja *et al.* (2021).

Continuando la línea abierta por estos y otros estudios, el objetivo de este trabajo es doble. Por un lado, exponemos los retos principales que deben ser enfrentados, a nuestro juicio, en el proceso de elaboración de corpus históricos que aporten datos fiables y adecuados para la aplicación de métodos cuantitativos en fonética histórica. Por otro lado, presentamos las características esenciales del corpus Oralía Diacrónica del Español (en adelante, ODE), prestando especial atención a la metodología que hemos seguido para la anotación de variantes gráficas de base fonética, y analizamos un estudio de caso con el objeto de demostrar la utilidad de este corpus para la investigación en fonética histórica del español clásico y moderno.

El trabajo se articula del modo siguiente: en el apartado 2 exponemos las características que a nuestro juicio debe cumplir un corpus histórico a fin de ofrecer documentación provechosa sobre la que extraer información de tipo fonético; en el apartado 3 explicamos la metodología que hemos adoptado en ODE para anotar cuatro variantes gráficas de base fonética: el seseo (*sapato* en lugar de *zapato*), el ceceo (*cilla* en lugar de *silla*), la variación en el uso de las consonantes líquidas (*flaire* en lugar de *fraile*) y la variación en el sistema vocálico (*tenaja* en lugar de *tinaja*); finalmente, cerramos este trabajo presentando en el apartado 4 un estudio de caso: el análisis de las grafías seseantes y ceceantes documentadas en los inventarios de bienes del siglo XVIII contenidos en ODE.

## 2. EL NIVEL GRÁFICO-FONÉTICO EN LOS CORPUS HISTÓRICOS: PROPUESTAS DE MEJORA

Es obvio que todo análisis en lingüística histórica ha de estar forzosamente basado en datos escritos. Partiendo de esta limitación, cabe preguntarse hasta qué punto es posible estudiar variantes fonéticas, habida cuenta de que todas las observaciones que hagamos sobre conductas de habla en el pasado no dejan de ser interpretaciones construidas sobre la variación ortográfica reflejada en la escritura y, en

concreto, sobre la que revelan las grafías de los textos que nos han llegado<sup>1</sup>. Por otro lado, es igualmente obvio que los corpus informatizados se han convertido en una herramienta esencial para el lingüista histórico, que encuentra en ellos un gran volumen de evidencia escrita sobre la que extraer de forma rápida numerosos ejemplos y frecuencias de uso. De hecho, no resulta exagerado afirmar que el corpus como herramienta de trabajo es el «*sine qua non* de la lingüística histórica» (McEnery y Wilson, 2001: 123). En consecuencia, la pregunta anterior puede reformularse en términos de utilidad de los corpus para el objetivo que aquí nos ocupa; es decir, cabe preguntarse acerca de los requisitos que debe reunir un corpus histórico para ofrecer la mejor solución posible ante la distancia, siempre inevitable, que separa la dimensión ortográfica de los textos de su repercusión e interpretación en el plano oral.

En este sentido, la idea de partida que adoptamos en el presente artículo es que la dificultad para explorar aspectos fonéticos en diacronía no es consustancial a los datos que manejamos en lingüística histórica —por fuerza, datos escritos—, sino que es atribuible a razones prácticas derivadas de las características de los corpus informatizados que tenemos a nuestra disposición.

The principles and guidelines of quantitative historical linguistics can be applied within any conventional area of historical linguistics, such as phonology, morphology, syntax, and semantics. This might at first glance seem to exclude e.g. historical phonology from quantitative historical linguistics; however, this is a practical consideration based on available corpus resources, not an inherent feature of quantitative historical linguistics. (Jenset y McGillivray, 2017: 36-37)

Es viable, por tanto, reunir un amplio conjunto de datos que puedan ser utilizados como base para la realización de análisis cuantitativos sobre fonética histórica. Para ello, contemplamos al menos tres aspectos que deben ser tenidos en cuenta en el proceso de construcción de un corpus histórico. Son los que enumeramos a continuación:

1. La recopilación de textos próximos a la oralidad y producidos por autores pertenecientes a diferentes zonas dialectales.
2. La elaboración de transcripciones basadas en criterios únicos de edición y en las que se respeten escrupulosamente las grafías de la fuente original.
3. La preparación de un nivel de edición, paralelo a la transcripción original, con grafías normalizadas según el estándar actual.

<sup>1</sup> Existen también fuentes de información indirectas para investigar el habla en las sociedades del pasado, sobre todo de la Edad Moderna, como son los comentarios realizados por gramáticos y lexicógrafos coetáneos, aunque estas fuentes no están exentas de problemas (Smith, 1997: 16).

El primer aspecto atiende al proceso de selección de datos y está pensado para garantizar la presencia de rasgos lingüísticos vernáculos, e idealmente geolocalizables, entre la documentación que va a integrar el corpus. El segundo aspecto atiende al proceso de edición de los materiales seleccionados y está encaminado a obtener transcripciones de calidad, fieles al texto original y libres de discrepancias internas. El último aspecto está relacionado con la fase de procesamiento de datos y tiene como finalidad aumentar la precisión en la extracción de información del corpus y mejorar sus posibilidades de explotación sobre el nivel gráfico-fonético.

Nótese que ninguna de las tres condiciones expuestas, que explicamos brevemente en los subapartados siguientes, está presente en los corpus históricos de referencia para el español. Una objeción formulada habitualmente al uso de estos corpus es la escasa presencia de textos representativos de la lengua hablada, así como de textos producidos por los estratos sociales más bajos, lo que dificulta el estudio de las variedades vernáculas empleadas en el pasado (Blas Arroyo, 2012: 1739-1740; Rodríguez Puente, 2018: 90-91). Además, el volumen de datos que exige un corpus de referencia obliga a echar mano de ediciones previamente elaboradas, cuyas particularidades no siempre coinciden con las que el lingüista idealmente necesita (Honkapohja *et al.*, 2009: 456-458). Y, al carecer de un conjunto sistemático de transcripciones fieles a las grafías de la fuente original, la estandarización ortográfica resulta a todas luces irrelevante, puesto que la obtención de variantes gráficas asociadas a una misma forma léxica no está garantizada en este tipo de corpus más generales.

## 2.1. Selección de textos próximos a la oralidad

Es sabido que al lingüista interesado en analizar datos históricos se le plantea una discrepancia evidente entre lo deseable —los datos con los que querría trabajar— y lo disponible —los datos con los que debe trabajar—, y que entre los factores que suscitan esta discrepancia destaca especialmente la carencia de fuentes directas de lengua hablada. La lingüística histórica en general y especialmente la pragmática y la sociolingüística históricas necesitan acceder a muestras de uso oral de la lengua, verdadero motor del cambio lingüístico, al tiempo que deben asumir la palabra escrita como fuente legítima de datos:

Tal situación ha llevado a la Lingüística histórica a un comportamiento que podríamos calificar de ‘esquizofrénico’: el único corpus que puede manejar directamente, al menos hasta el momento actual, es el escrito [...], pero el lingüista suele renegar de él e intenta hallar en lo escrito algo que le remita más allá, a un mundo comunicativo que se le oculta y del que la escritura apenas le revela leves indicios. (Cano Aguilar, 1996: 376)

Con el propósito de sortear este problema, desde los años noventa del siglo pasado los estudios histórico-lingüísticos han comenzado a poner el foco de atención en diferentes documentos escritos que se aproximan al polo de la oralidad, ya sea por tratarse de textos basados en un acto real de lengua hablada o por contener rasgos frecuentes de lengua hablada o porque fueron diseñados para ser articulados de forma oral (Culpeper y Kytö, 2010: 18). Nos referimos a cartas privadas, crónicas de soldados, peticiones de ayuda a la beneficencia, declaraciones de testigos, actas inquisitoriales, inventarios de bienes o sermones, entre otras tipologías. Todos estos casos comparten la particularidad de ser documentación caracterizada por su inmediatez comunicativa, en mayor o menor grado (Oesterreicher, 2004).

Aún más interesante para los propósitos de este trabajo es el hecho de que este cambio de foco también se haya visto reflejado en la propia compilación de corpus históricos. Así, en los últimos años, y sin salir del ámbito hispánico, se han emprendido proyectos destinados a la elaboración de corpus y bases de datos formados exclusivamente por tipos textuales de inmediatez comunicativa. En este sentido, cabe citar al menos tres ejemplos paradigmáticos: el corpus P. S. Post Scriptum (Vaamonde, 2018), formado por cartas privadas escritas en Portugal y España entre 1500 y 1830 y con un tamaño que ronda los dos millones de palabras (un millón por cada lengua: portugués y español); el corpus CorLexIn: Corpus Léxico de Inventarios (Morala Rodríguez, 2018a, 2018b), constituido por inventarios de bienes de la época áurea y con un tamaño actual de 1 800 000 palabras; y el corpus ODE: Oralia Diacrónica del Español (Calderón Campos y Vaamonde, 2020), constituido en la actualidad por algo de más de un millón de palabras extraídas de declaraciones de testigo e inventarios de bienes, fundamentalmente. A ellos podríamos sumar también el subcorpus constituido por cartas privadas y por «notas y breves» que incluye CODEA+2022.

La consulta de estos recursos digitales, todos ellos de acceso libre en la red, permite extraer con cierta facilidad estructuras lingüísticas que se desvían de la tradición normativa y que dejan aflorar fenómenos de variación propios del subestándar, correspondientes a diferentes niveles de análisis y no siempre atestiguados en los corpus generales: casos de dequeísmo (1a), adverbios no estandarizados (1b), vocabulario dialectal (1c) o el reflejo en la escritura de pronunciaciones seseantes (1d):

- (1) a. *crei de que* me describirías (Post Scriptum, PSCR6538)
- b. *esta solo sirve para decirte ande as ido* (Post Scriptum, PS6229)
- c. *unas tenazas, una rasera que no tiene pala, un bail* (ODE, GR1829I2005)
- d. *unos sapatos de beserro morado en quinse rreales* (ODE, SE1780I7048)

Entre ellos, en este trabajo cobran especial importancia los aspectos que tienen que ver con cuestiones de tipo fonético (1d), y que se manifiestan con especial intensidad en este tipo de documentación oralizante en la que no existe un excesivo control por la forma.

## 2.2. Elaboración de ediciones diplomáticas

Evidentemente, para poder identificar casos como los recogidos en el último ejemplo de la lista anterior es indispensable contar con transcripciones exentas de cualquier tipo de intervención editorial en el nivel ortográfico. Reunir un conjunto amplio de datos en que se respete esta condición no es, sin embargo, tarea fácil. De hecho, los corpus históricos no siempre ofrecen este tipo de ediciones textuales, pues ello depende, en último término, del propósito para el que fueron construidos.

Los corpus históricos de referencia (CORDE, CdE-hist y CDH) están diseñados para dar cabida a un número amplio y variado de textos que aseguren la representatividad y faciliten el conocimiento global de una lengua en diferentes períodos del idioma. Deben priorizar, por tanto, el tamaño, y se valen para ello de ediciones ya hechas de los textos seleccionados, «con lo que ello supone de aparición de discrepancias en el sistema de edición seguido y en la calidad de las ediciones utilizadas» (Rojo, 2021: 78). Repárese en que este problema es, en realidad, consustancial a los corpus históricos generales: por su propia finalidad y por las características con que están diseñados, estos corpus son siempre fruto de «un trabajo de segundo nivel, que vive de la integración, adaptada, de lo que otros han hecho previamente» (Rojo, 2010: 1157). La alternativa a esta metodología —siempre en el terreno de la diacronía— consiste en priorizar el control filológico, reduciendo ostensiblemente el volumen de datos a cambio de ofrecer ediciones muy cuidadas que reproducen con exactitud la lengua del texto original, lo que redunda en beneficio de su posterior explotación lingüística.

Dollinger (2004) alude a esta disyuntiva a la que se enfrenta el compilador de corpus históricos con las expresiones *philological outsourcing* y *philological computing*, respectivamente<sup>2</sup>. En el primer caso, el compilador confía en el trabajo hecho previamente por el filólogo para seleccionar los textos que formarán parte del corpus: se ahorra tiempo y esfuerzo, pero la calidad y utilidad del corpus dependerá de la calidad de sus ediciones, así como de los principios editoriales adoptados. En el segundo caso, se crean ediciones propias para confeccionar el corpus: así se asegura una mayor homogeneidad y rigor filológico, aunque la transcripción de fuentes primarias es un proceso lento y costoso. Por eso, esta segunda vía da como resultado la creación de corpus históricos especializados, de pequeño tamaño, que en el ámbito hispánico representan, por ejemplo, Biblia Medieval (Enrique-Arias, 2008), el Corpus de Documentos Españoles Anteriores a 1900 (CODEA), el corpus P. S. Post Scriptum, el corpus CorLexIn, el Corpus Mallorca (s. f.) o el corpus Oralia Diacrónica del Español (ODE), entre otros.

En lo que atañe a la identificación de fenómenos de tipo fonético, la metodología adecuada solo puede ser esta última (*philological computing*), al menos por lo que se refiere a la ortografía del texto original. Cualquier intervención en este nivel

<sup>2</sup> Este último concepto fue propuesto inicialmente por Meurman-Solin (2001: 18).

de análisis, y en especial si no está debidamente documentada en el propio corpus, supone una adulteración del texto que reducirá las posibilidades de extraer de él ejemplos fiables de variación fonética:

If as is often the case, texts are ‘normalized’ (or as Penzl 1991 puts it ‘redialectalized’), this irons out variation, which yields a spurious appearance of formal homogeneity. And this may produce a kind of object that could not in fact have been written by anybody in the tradition the text comes from. (Lass, 1997: 101)

### 2.3. Normalización ortográfica del corpus

La cita anterior recoge una crítica a las ediciones normalizadas como sustitución de las transcripciones originales y, por tanto, a su uso como única vía de acceso a las fuentes históricas para su estudio lingüístico. El problema de esta práctica es que los datos originales (*first-order witnesses*) acaban siendo falseados y convertidos en constructos artificiales (*second-order witnesses*) (Lass, 1997: 100). Una cuestión diferente, al menos en lo que concierne al proceso de construcción y explotación de corpus históricos, es la elaboración de versiones normalizadas de los textos sin menoscabo de sus correspondientes transcripciones diplomáticas, esto es, el hecho de agregar a las formas ortográficas atestiguadas en la fuente original las formas ortográficas estandarizadas, palabra a palabra, para ofrecer una edición alternativa que complemente —no que sustituya— a la edición más conservadora.

Para comprender cabalmente los beneficios que aporta esta tarea de normalización a un corpus histórico, debemos tener presente la alta variación ortográfica que suelen presentar las fuentes históricas, más notoria si cabe cuanto mayor sea la profundidad diacrónica del corpus. De la palabra *vecino*, por ejemplo, se documentan en ODE hasta trece variantes gráficas diferentes al momento de redactar estas líneas (*vezino*, *vecino*, *veçino*, *vesino*, *uezino*, *bezino*, *becino*, *ueçino*, *beçino*, *uecino*, *vessino*, *besino*, *uesino*), y eso sin contar las numerosas abreviaturas atestiguadas para referirse a este mismo concepto (*vezo*, *vo*, *vzo*, *vso*, *uo*, *vez*, *bo*, *veo*, *bzo*...).

Como resulta fácil de imaginar, esta variación ortográfica constituye un problema importante a la hora de trabajar con un corpus histórico, tanto para extraer información como para realizar tareas de procesamiento automático. Contar, en cambio, con una versión normalizada permite al usuario recuperar todos los ejemplos de una voz o de una expresión determinadas con independencia de su escritura (por ejemplo, la obtención de todas las ocurrencias de la palabra *vecino*), y mejora el resultado de las herramientas de etiquetado automático, al poder ser entrenadas sobre un conjunto de datos mucho más homogéneo (por ejemplo, un anotador automático solo debe etiquetar correctamente la palabra *vecino*, y no todas sus variantes ortográficas):

The argument for normalisation is twofold. First, that it helps to improve the accuracy of automated computational linguistic (natural language processing) techniques such as part-of-speech tagging and second, that it improves the stability and robustness of corpus linguistic methods such as keyword analysis, thereby allowing existing software tools of both types to be used unmodified. (Archer *et al.*, 2015: 6)

A estas dos ventajas cabe sumar, además, una tercera, no indicada en la cita anterior y sobre la que queremos llamar la atención en este trabajo. Nos referimos a la posibilidad de contrastar las formas ortográficas originales con sus correspondientes formas normalizadas con el objetivo de identificar y anotar automáticamente aquellas variantes gráficas que presenten un especial interés en el plano fonético de la lengua. En otras palabras, a partir de la comparación de ambas versiones, formas originales y formas normalizadas, es posible detectar diferencias ortográficas con base fonética, como pueden ser casos de seseo (*sinco* frente a *cinco*) y ceceo (*cauze* frente a *causa*), así como cambios relacionados con el vocalismo átono (*sepoltura* frente a *sepultura*) o con la posición de las consonantes líquidas (*corchon* frente a *colchón*). De cómo se ha implementado esta tarea en ODE y de los resultados obtenidos se ocupa el siguiente apartado.

### 3. CORPUS Y METODOLOGÍA

El corpus ODE es un corpus histórico especializado en donde confluyen las tres condiciones expuestas en el apartado anterior. En primer lugar, se trata de un corpus compuesto fundamentalmente por dos tipos textuales próximos al plano de lo oral y a la esfera de lo cotidiano: las declaraciones de testigo, en donde el escribano deja constancia por escrito de los testimonios que oye, y los inventarios de bienes, en donde el escribano enumera de forma minuciosa una lista de bienes. En ambos casos, el contenido textual se genera de forma similar: el declarante habla y el escribano apunta lo que oye con el único objetivo de reflejar fielmente la verdad de los hechos, ya sea esta el testimonio de un testigo —ocasionalmente, en estilo directo— o la descripción y tasación de unos enseres, que generalmente se refieren a objetos de la vida cotidiana. Además, tanto los informantes que declaran como los escribanos que anotan proceden, por lo general, del área geográfica donde se han producido los hechos documentados, que siempre aparece recogida en el manuscrito. Esta particularidad tiene especial relevancia, puesto que abre la posibilidad de proyectar los datos del corpus en un mapa para investigar el alcance dialectal de un determinado fenómeno, como se verá en el apartado siguiente.

En segundo lugar, todos los textos que conforman el corpus ODE han sido transcritos siguiendo unos mismos criterios de edición, que fueron planteados con la intención de crear transcripciones muy conservadoras. Se han respetado las grafías originales, así como el empleo de abreviaturas. También se han marcado aspectos de índole

paleográfica como las adiciones fuera de línea, las cancelaciones o las lagunas textuales, junto con otros aspectos editoriales como conjeturas, erratas o lecturas dudosas. En ODE, toda esta información está almacenada y estructurada en lenguaje XML y siguiendo los estándares propuestos por el consorcio TEI (*Text Encoding Initiative*) para la representación de fuentes primarias en formato digital (Burnard, 2014).

En tercer y último lugar, el corpus ODE ofrece tres vías de acceso al documento: la transcripción conservadora, la imagen facsimilar y una versión con ortografía normalizada según las normas del español estándar actual. Para la elaboración de esta última se ha hecho uso de la plataforma TEITOK (Janssen, 2016), que incluye un normalizador automático diseñado para poder usar los propios datos, una vez normalizados, como corpus de entrenamiento; de esta forma, los resultados de la herramienta mejoran a medida que aumenta el tamaño del corpus. La plataforma TEITOK incluye asimismo otro tipo de funcionalidades de las que se vale el corpus ODE para mejorar su rendimiento y utilidad, como son la visualización de los datos en la web, el sistema de edición del corpus, la anotación morfosintáctica o la propia interfaz de consulta (Calderón Campos, 2019).

Las tareas de procesamiento lingüístico aplicadas en el corpus ODE se ejecutan de manera secuencial a través de la plataforma TEITOK. El punto de partida en este proceso es el conjunto de ediciones conservadoras en lenguaje XML-TEI. Sobre el contenido de esta edición, se delimita cada palabra y signo de puntuación mediante un proceso de *tokenización* y, a continuación, a cada unidad identificada en el texto (*token*) se le van añadiendo diferentes pares de atributo-valor que se corresponden con diferentes niveles de edición textual: desarrollo de abreviaturas (si fuera el caso), normalización ortográfica, etiquetación morfosintáctica y lematización. Por ejemplo, el resultado final de la palabra *bezinos*, una vez realizadas estas tareas de procesamiento lingüístico, es el que se muestra en la Figura 1; para el caso de la abreviatura *bzos*, el resultado se muestra en la Figura 2<sup>3</sup>:

**Figura 1**

*Marcación de la forma bezinos en ODE*

```
<tok nform="vecinos"
      pos="NCMP000"
      lemma="vecino">bezinos</tok>
```

*Nota.* Fuente: elaboración propia

<sup>3</sup> tok = token; fform = full form ('forma expandida'); nform = normalized form; pos = part of speech; NCMPooo = Nombre Común Masculino Plural.

**Figura 2***Marcación de la forma bzos en ODE*

```
<tok fform="bezinos"
      nform="vecinos"
      pos="NCMP000"
      lemma="vecino">bzos</tok>
```

*Nota.* Fuente: elaboración propia

Los dos niveles que nos interesa comparar en este trabajo son la forma original, representada como contenido del elemento `<tok>` (*bezinos* en la Figura 1) y la forma normalizada, representada como valor del atributo `@nform` (*vecinos* en la Figura 1). Si la forma original es una abreviatura, entonces se toma en consideración para la comparación la forma expandida correspondiente, representada como valor del atributo `@fform` (*bezinos* en la Figura 2). Mediante la comparación de los valores correspondientes a estos niveles de edición se pueden detectar fenómenos fonéticos como los que se recogen en la Tabla 1:

**Tabla 1***Forma original, forma normalizada y fenómeno fonético identificado*

Forma original	Forma normalizada	Fenómeno fonético
<i>lienzo, vesino</i>	<i>lienzو, vecino</i>	seseo
<i>carmezi, cilla</i>	<i>carmesí, silla</i>	ceceo
<i>arjofar, cobertol</i>	<i>aljófar, cobertor</i>	consonantes líquidas
<i>sepoltura, defunto</i>	<i>sepultura, difunto</i>	cambios vocálicos

*Nota.* Fuente: elaboración propia

Este tipo de fenómenos han sido identificados y anotados de forma automática mediante la ejecución de un *script* escrito en lenguaje Perl. El algoritmo utilizado para este fin consiste en efectuar una comparación, letra a letra, de la forma original y de la forma normalizada de cada *token* (o de la forma expandida y de la forma normalizada, si se trata de abreviaturas); a continuación, el *script* contrasta el resultado de la comparación con una lista preestablecida de correspondencias, que están relacionadas con los cuatro fenómenos recogidos en la Tabla 1; finalmente, y en caso de que haya un resultado positivo —es decir, alguna coincidencia con la lista de correspondencias—, se anota el fenómeno fonético pertinente en un nuevo atributo, denominado `@phtags` (*phonetic tags*).

Por ejemplo, al comparar la forma original *vesino* con la forma normalizada *vecino*, el *script* identifica que en la tercera posición de esa palabra la letra *s* fue normalizada a una *c* seguida de una *e*; puesto que esa correspondencia *s-c(e)* está asociada al fenómeno del seseo, el *script* añade automáticamente el atributo @phtags con el valor de «seseo», como se muestra en la Figura 3:

**Figura 3**

*Marcación de la forma vesino en ODE/TEITOK*

```
<tok nform="vecino"
      pos="NCMS000"
      lemma="vecino"
      phtags="seseo">vesino</tok>
```

*Nota.* Fuente: elaboración propia

Entre las correspondencias que se han tenido en cuenta para la anotación de variantes ortográficas de base fonética, destacan las que recogemos en la Tabla 2. Todos los ejemplos recogidos en esta tabla están documentados en el corpus ODE. Cabe señalar que el valor escogido para la anotación de algunos de estos fenómenos tiene un marcado carácter general. Por ejemplo, el valor «liquid\_consonants» comprende diferentes fenómenos relacionados que pueden ser recuperados mediante una sola consulta, dejando a juicio del usuario si se trata de casos de lambdacismo, rotacismo, metátesis, asimilación, disimilación, etcétera.

El *script* que se ha ejecutado también ha tenido en cuenta la posibilidad de que una misma palabra ilustre dos o más fenómenos simultáneamente. Por ejemplo, en ODE se documenta la forma *balseletes*, normalizada a *brazaletes*, y cuyo atributo @phtags ha sido anotado con los valores de «seseo», «vowel\_system» y «liquid\_consonants».

Finalmente, y como cabría esperar, este proceso de anotación automático demandó la realización de campañas de revisión manual, con el propósito de eliminar algunos falsos positivos que fueron anotados erróneamente por el *script*. Por ejemplo, las formas del tipo *parescer* y similares fueron anotadas automáticamente como casos de seseo, debido a la correspondencia *s-c(e)* producida en la quinta posición de esta palabra (*parescer*, *parecer*). Al momento de redactar estas líneas, este tipo de errores ya han sido corregidos.

Actualmente, en el corpus ODE se documentan 6728 formas anotadas como «seseo», 1211 formas anotadas como «ceceo», 1412 formas anotadas como «liquid\_consonants» y 4148 formas anotadas como «vowel\_system». Este conjunto de datos, sumado a la posibilidad de controlar la procedencia geográfica de cada documento con un alto nivel de certeza, convierte a ODE en un recurso digital idóneo para realizar estudios cuantitativos sobre cualquiera de estos cuatro fenómenos fonéticos.

Tabla 2

Variantes gráficas consideradas para la anotación de fenómenos fonéticos en ODE

Forma original	Forma normalizada	Correspondencia	Valor de @phtags
<i>calsones, posso</i>	<i>calzones, pozo</i>	s–z, ss–z	seseo
<i>calsetas, acaesido</i>	<i>calcetas, acaecido</i>	s–c(e), s–c(i)	seseo
<i>besserro, vessino</i>	<i>becerro, vecino</i>	ss–c(e), ss–c(i)	seseo
<i>amazar, çarten</i>	<i>amasar, sartén</i>	z–s, ç–s	ceceo
<i>bracero, cilla</i>	<i>brasero, silla</i>	c(e)–s, c(i)–s	ceceo
<i>tabulete, albañyr</i>	<i>taburete, albañil</i>	l–r, r–l	liquid_consonants
<i>flaire, corgal</i>	<i>fraile, colgar</i>	l_r–r_l, r_l–l_r	liquid_consonants
<i>esturche, albellanas</i>	<i>estuche, avellanas</i>	r–Ø, l–Ø	liquid_consonants
<i>moselina, duzientos</i>	<i>muselina, doscientos</i>	o–u, u–o	vowel_system
<i>ligitima, tenaja</i>	<i>legítima, tinaja</i>	i–e, e–i	vowel_system
<i>monesterio, rodopies</i>	<i>monasterio, rodapiés</i>	e–a, o–a	vowel_system
<i>tabarneros, ascuras</i>	<i>taberneros, oscuras</i>	a–e, a–o	vowel_system

Nota. Fuente: elaboración propia

#### 4. SESEO Y CECEO EN LOS INVENTARIOS DE BIENES DEL SIGLO XVIII DE ODE

Para este estudio de caso se van a tener en cuenta, exclusivamente, los inventarios de bienes del siglo XVIII del corpus ODE, cuyo tamaño se cuantifica en la Tabla 3. En el subcorpus del XVIII están representadas hasta la fecha seis provincias andaluzas, dos orientales (Granada y Almería) y cuatro occidentales (Huelva, Cádiz, Sevilla y Málaga). Se ha incluido, además, la provincia de Madrid, cuyos datos servirán de corpus de control, imprescindible para interpretar adecuadamente el valor fonético de las grafías «supuestamente» seseantes o ceceantes de las provincias andaluzas.

En este apartado se persiguen dos objetivos. El primero, de carácter general, se relaciona con el valor de las grafías indicadoras de una posible neutralización de /s/ y /θ/: «s» en lugar de «z» o «c» (*alcusa, sinco*), y «z», «c» o «ç» en lugar de «ss» (*yglezia, ciituada, conçentimiento*)<sup>4</sup>. Se trata de determinar si estas facilidades grafemáticas pueden interpretarse con valor fonético, o si por el contrario son meras confusiones ortográficas casuales. Si la respuesta a esta primera pregunta se decanta a favor de que la variación gráfica tiene base fonética, el segundo objetivo es determinar el

<sup>4</sup> Un estado de la cuestión reciente sobre la historia de las sibilantes en español puede verse en Núñez-Méndez (2021).

alcance del seseo/ceceo en las tres regiones en que se ha dividido el subcorpus de inventarios del siglo XVIII: Andalucía occidental, Andalucía oriental y Madrid<sup>5</sup>.

Tabla 3

*Subcorpus de inventarios de bienes del siglo XVIII de ODE considerados en este artículo*

Provincia	N.º de tokens
Almería	32 018
Granada	52 197
Cádiz	69 844
Huelva	28 388
Málaga	12 833
Sevilla	32 877
Madrid	34 181

Nota. Fuente: elaboración propia

#### 4.1. ¿Errores ortográficos casuales o indicios de seseo/ceceo?

En un testamento de Huelva de 1709 (ODE HU1709Io325)<sup>6</sup> puede leerse, en versión paleográfica, lo siguiente:

(2) Yten declaro me deue Juana Baptista, mi suegra, tres fanegas de trigo, las dos que le preste el año pasado de *setessientos* y ocho, a *presio* de *sinquenta* reales faneña, y la otra que pagó de mi caudal Bartolomé Barrera, mi marido difunto [...] y assimismo sesenta y tres reales que pagué por la *suzodicha* de veinte y una peonada<sup>7</sup> en sus biñas. (ODE HU1709Io325)

Se han marcado en cursiva los indicios de neutralización fonológica (*setessientos*, *presio*, *sinquenta* y *suzodicha*) del ejemplo (2), que son muy frecuentes en el testamento onubense: *Consepción* (5 casos), *grasias*, *albaseas*, *intension*, *deuosiones*, *asseite*, *sinco*, *poso*, *dulse*, *presioza*, *yglezia*, *miza*, *pazado*, *cazas* ‘casas’, etc.<sup>8</sup>.

<sup>5</sup> Para un estudio del seseo en el siglo XVIII, en cada provincia andaluza, véase González Sopena (2023: 153-156).

<sup>6</sup> Puede verse el inventario en el siguiente enlace: <http://corpora.ugr.es/ode/index.%20php?action=file&cid=ode/corpus/HU1709Io325.xml>

<sup>7</sup> Obsérvese la ausencia de -s en ‘veinte y una *peonada*», y más adelante en el mismo testamento en ‘un barril de sardina», probablemente reflejos de la relajación de la -s implosiva. Este aspecto fonético es muy difícil de rastrear cuantitativamente en la documentación.

<sup>8</sup> A estos ejemplos seseantes/ceceantes, pueden añadirse confusiones -r/-l (*engalsado*, *alcos* ‘arcos’) o yeístas (*xolla* ‘joya’).

Cuando se producen «confusiones» ortográficas de este tipo (es decir, *asseite* por *aceite* o *miza* por *misa*, etc.) la discusión se centra en determinar si se trata de vacilaciones ortográficas casuales o si, por el contrario, pueden interpretarse como pruebas de desfonologización /s/ y /θ/.

En este apartado vamos a intentar demostrar que existe una diferencia estadísticamente significativa en la frecuencia de «s» por «z» o «c» (y viceversa) en los inventarios de las tres regiones analizadas. Si la confusión fuera puramente ortográfica (sin bases fonéticas) habría que esperar porcentajes similares de vacilación en los escribanos madrileños y andaluces, y entre los de la Andalucía occidental y los de la oriental. Si, por el contrario, en los inventarios de algunas de las regiones estudiadas se producen porcentajes de alternancias ortográficas significativamente diferentes a los de otras regiones, habrá que considerar que tal variación se relaciona con el origen geográfico del documento y que, en consecuencia, tiene motivación fonética, seseante o ceceante.

Somos conscientes de que la etiqueta de ceceo puede no corresponderse con casos auténticamente ceceantes. El sentido fonético real (seseo o ceceo) dependerá del periodo histórico y de la región en la que se produzca la vacilación ortográfica: variantes ceceantes en documentación onubense del siglo XVIII o XIX tendrán más posibilidades de interpretarse como auténticos casos de ceceo que si se tratara del mismo fenómeno en documentación de la capital mexicana de la misma época, debido al hecho evidente de que los atlas lingüísticos del siglo XX registran abundantes casos de ceceo en Huelva, pero no en México.

En cualquier caso, las etiquetas «ceceo» y «seseo» nos permiten separar las dos direcciones de las neutralizaciones ortográficas (sin salir, insistimos, de lo estrictamente ortográfico): a favor de «c», «ç» o «z» (ceceo) o a favor de «s» (seseo). La interpretación fonética deberá hacerse a partir del conocimiento dialectal de la zona. Además, no debe olvidarse que su interpretación como simple error ortográfico o como reflejo de la oralidad es lo que se está discutiendo en este párrafo.

Como se ha dicho en § 3, disponer en ODE de formas paleográficas (*form*) y poder compararlas automáticamente con sus correspondientes formas normalizadas (*nform*) genera la enorme ventaja metodológica de disponer, mediante una búsqueda muy sencilla, de todos los casos etiquetados como «seseo» o «ceceo» de una época o de una región determinada (ver Tabla 4).

Queremos determinar si existe una relación estadísticamente significativa entre procedencia del inventario y número de «confusiones» ortográficas interpretables como seseo/ceceo. Como paso previo:

- a. Se han establecido tres subcorpus regionales del siglo XVIII que sirvan de contraste entre sí: Andalucía occidental, Andalucía oriental y Madrid.
- b. Se han identificado todos los casos de discrepancia entre «s», por un lado, y «c», «ç» o «z», por otro, comparando la forma paleográfica y la normalizada.

Todos los casos se han etiquetado como *phtags* = “seseo” o *phtags* = “ceceo”, según se ha explicado en § 3.

Tabla 4

Diez casos más frecuentes de seseo y ceceo en inventarios onubenses del siglo XVIII (ODE)

Ceceo	N.º ejemplos	Seseo	N.º ejemplos
<i>açi</i>	7	<i>sinco</i>	125
<i>quezada</i>	5	<i>dies</i>	53
<i>Luiza</i>	5	<i>quinse</i>	32
<i>citio</i>	4	<i>sinuenta</i>	31
<i>cauza</i>	4	<i>apresiado</i>	25
<i>çu</i>	3	<i>siento</i> ('ciento')	20
<i>suzodicha</i>	3	<i>Dias</i> ('Díaz')	20
<i>yglecia</i>	2	<i>resiuo</i>	19
<i>recidente</i>	2	<i>apresiada</i>	16
<i>fueze</i>	2	<i>apresiados</i>	16

Nota. Fuente: elaboración propia

Como es habitual, se parte de una hipótesis nula, según la cual no hay diferencia significativa en la proporción de palabras con seseo/ceceo entre los tres corpus, es decir, cualquier variación observada entre los subcorpora podrá atribuirse al azar.

Para rechazar la hipótesis nula se ha realizado un test de Chi<sup>2</sup>, habitualmente usado en estadística para evaluar si existe una relación significativa entre dos variables categóricas, en nuestro caso, lugar de origen del inventario y presencia (o no) de confusiones gráficas relacionadas con el seseo/ceceo. Básicamente, el test de Chi<sup>2</sup> lo que hace es comparar las frecuencias observadas en los datos con las frecuencias que se esperarían si no hubiera relación entre las variables (hipótesis nula).

Una manera muy sencilla y fiable de realizar esta prueba de independencia es hacerlo a través de la web de la Universidad de Lancaster (<http://corpora.lancs.ac.uk/sigtest/>): el sistema utiliza R de fondo, pero los usuarios solo tienen que aportar la frecuencia absoluta del fenómeno que quieran estudiar y el tamaño de los diferentes subcorpora. En nuestro caso, basta con proporcionar los datos que se indican en la Tabla 5.

Con estos datos se obtienen los resultados de la Tabla 6. El valor de p es prácticamente 0, lo que debe interpretarse como una muy fuerte evidencia contra la

hipótesis nula<sup>9</sup>. En nuestro contexto de estudio, este dato significa que las diferencias observadas en el uso de las grafías «s» / «c», «ç», «z» en las tres regiones hispánicas no se deben al azar, sino que se relacionan con el lugar de escritura del inventario. Por tanto, parece más razonable interpretar estas confusiones gráficas no como meros errores ortográficos sino como reflejo del seseo/ceceo de la zona donde esas confusiones tienen un mayor índice.

Tabla 5

*Casos de seseo/ceceo en inventarios del XVIII en ODE*

	Seseo/ceceo	Tamaño del corpus en tokens
Andalucía occidental	3678	143 942
Andalucía oriental	210	84 135
Madrid	25	34 181

*Nota.* Fuente: elaboración propia

Tabla 6

*Frecuencias absolutas y esperadas de grafías seseantes/ceceantes en inventarios de los tres subcorpus del XVIII*

Región	Frec. observada seseo/ceceo	Frec. esperada seseo/ceceo
Andalucía occidental	3678	2170
Andalucía oriental	210	1239
Madrid	25	502

$\chi^2 = 2392.44$ ,  $df = 2$ ,  $p < 0.001$

*Nota.* Fuente: elaboración propia

La comparación de las frecuencias absolutas con las esperadas (Tabla 6) ayuda a entender el panorama general que plantean los datos: en Andalucía occidental la frecuencia observada del seseo/ceceo es mucho más alta que las esperadas; por el contrario, las frecuencias observadas de Andalucía oriental y Madrid están muy por debajo de las expectativas. Esto significa que, en el siglo XVIII, la neutralización de s/θ había avanzado en el occidente andaluz mucho más que en la zona oriental de Andalucía y, por supuesto, mucho más que en Madrid.

<sup>9</sup> En la misma página, se obtiene un valor de G (logaritmo de verosimilitud o Log Likelihood) de 3020, lo que confirma un valor de p próximo a 0.

#### 4.2. Las dos Andalucías lingüísticas en el siglo XVIII

La discusión de § 4.1 se ha centrado en demostrar que las diferencias gráficas observadas en los tres subcorpus tienen una base fonética bastante clara. En este apartado vamos a cartografiar los datos de seseo/ceceo disponibles hasta la fecha (febrero de 2024) en ODE. Debido a que la red de localidades todavía no es muy extensa (2 o 3 puntos por provincia), hemos optado por una representación provincial, que nos dará una visión aproximada de la extensión dieciochesca de la neutralización de s/θ.

El cartografiado se ha realizado en R, con las bibliotecas «mapSpain» (Hernández, 2024) y ggplot2, teniendo en cuenta la frecuencia por millón de grafías etiquetadas como seseo o ceceo en ODE, para cada una de las seis provincias andaluzas actualmente disponibles (véase la Tabla 7).

Tabla 7

*Frecuencias por millón de casos de seseo/ceceo en 6 provincias andaluzas (siglo XVIII, ODE)*

Provincia	Seseo/ceceo por millón
Almería	2748
Granada	2452
Cádiz	18 054
Huelva	36 218
Málaga	31 436
Sevilla	31 359

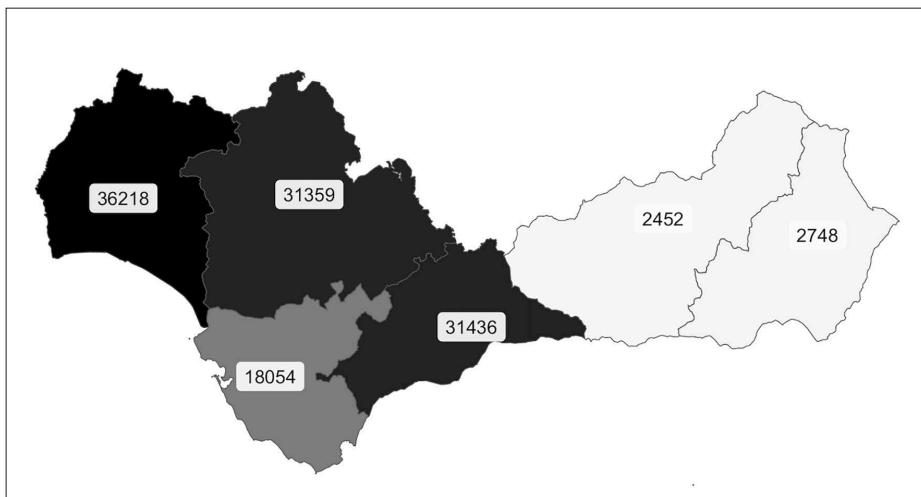
Nota. Fuente: elaboración propia

El mapa de la Figura 4 permite ver claramente la división que se establecía en el siglo XVIII entre el occidente y el oriente andaluz. La región occidental tiene una media ligeramente superior a 29 000 casos de seseo/ceceo por millón, que desciende en el oriente a 2600. Podría decirse, para entender mejor la diferencia, que en el subcorpus de inventarios del occidente es once veces más probable encontrar casos de seseo/ceceo que en el corpus oriental.

Por último, resulta muy significativa la comparación del mapa del siglo XVIII con el que se muestra en la Figura 5 realizado a partir de datos del ALEA o *Atlas lingüístico y etnográfico de Andalucía* (mapa 1705) proporcionados por Frajeda Rueda (2019), que aplica una interpolación de Voronoi a los puntos de encuesta del atlas. Las áreas de neutralización (seseo, ceceo, polimorfismo) del atlas de Andalucía se concentran en las mismas regiones occidentales detectadas dos siglos antes en el corpus ODE.

Figura 4

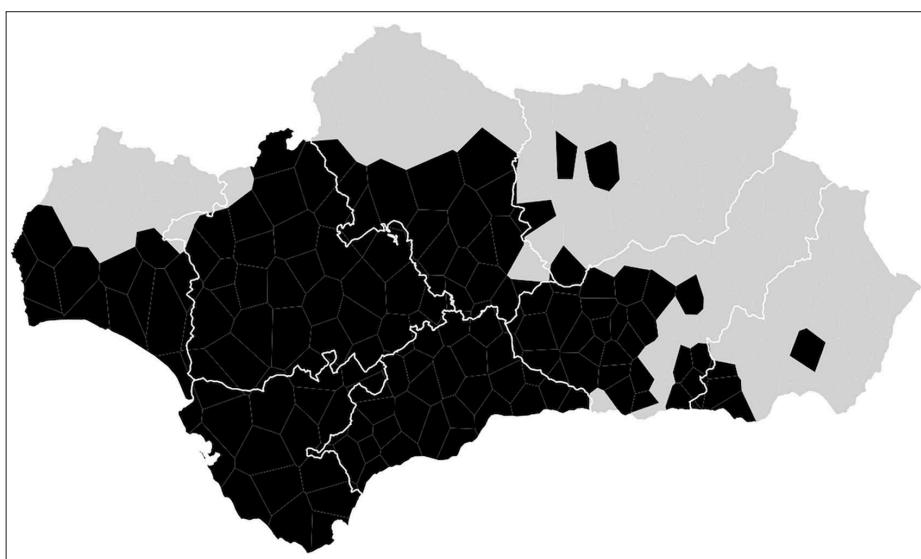
Expansión del seseo/ceceo por Andalucía en el siglo XVIII, según datos de ODE



Nota. Fuente: elaboración propia

Figura 5

Neutralización (en negro) y distinción (gris) de s/θ según datos del ALEA (1961-1973)



Nota. Fuente: elaboración propia a partir de los datos de Fradejas Rueda (2019)

## 5. CONCLUSIONES

Se ha querido mostrar en este artículo que la estrategia de TEITOK, en la que en cada *token* se vincula una forma original o paleográfica (*form*) con una normalizada (*nform*), convierte a ODE en una herramienta de enorme valor para analizar fenómenos fonéticos en perspectiva histórica. Gracias a ello, mediante el *script* que compara ambas formas (*suzodicha/susodicha*, *presio/precio*), es posible cuantificar la variación gráfica y determinar su valor fonético. En el estudio de caso se ha podido comprobar que las vacilaciones de sibilantes presentes en ODE («s» / «c», «ç», «z») no parecen deberse al azar, puesto que en tal caso la proporción de vacilaciones sería equivalente en todas las regiones del corpus (hipótesis nula). El análisis estadístico ha permitido comprobar una frecuencia significativamente decreciente del seseo y ceceo en Andalucía occidental, oriental y Madrid, lo que dibuja con claridad dos contrastes dialectales en el seseo/ceceo del XVIII: uno general, que diferencia a Andalucía de Madrid; y otro intrarregional, que divide Andalucía entre el occidente y el oriente.

Por otra parte, se ha puesto el énfasis en las ventajas metodológicas que ofrece la comparación entre forma paleográfica y forma normalizada, pero ha servido también para reconocer que los avances metodológicos (buscadores avanzados, métodos estadísticos, scripts de comparación, etc.) necesitan de un trabajo filológico previo e intenso. El análisis del seseo/ceceo en el siglo XVIII ha servido para indicar las líneas filológicas futuras de ampliación del corpus ODE, que tendrán que centrarse en los dos aspectos siguientes:

- a. Ampliación provincial (es urgente, en este caso, disponer de datos de Córdoba y Jaén) y aumento del número de puntos disponibles por provincia, para poder generar mapas que no estén encorsetados por los límites administrativos.
- b. Ampliación cronológica, con documentación de los siglos XVI y XVII, para cartografiar el avance del seseo/ceceo desde el origen de la Edad Moderna.

Por último, dado que cada vez son más los grupos de investigación dedicados a la recuperación del patrimonio documental hispánico, se hacen muy necesarias labores de coordinación y de fijación de criterios, como la emprendida por la red internacional CHARTA, cuyos principios básicos se asumen en nuestro proyecto de corpus ODE.

## FINANCIACIÓN

Esta publicación es parte del Proyecto PID2022-136256NB-I00 financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

## CONTRIBUCIONES DE AUTORÍA

Este artículo ha sido escrito por dos autores. El criterio para el orden de firma es puramente alfabético. Miguel Calderón Campos ha contribuido realizando el estudio de caso (apartado 4) y las conclusiones, mientras que Gael Vaamonde ha contribuido realizando la introducción, el estado de la cuestión y la metodología (apartados 1, 2 y 3). La bibliografía y la revisión global del artículo se han realizado conjuntamente.

## REFERENCIAS

Almeida Cabrejas, B. (Coord.). (s. f.). *Corpus hispánico y americano en la red: textos antiguos (CHARTA)*. <https://www.corpuscharta.es/consultas.html>

Almeida Cabrejas, B. (2014). Scriptores con bajo y medio nivel socioeducacional en documentos del siglo XIX del Archivo Municipal de Alcalá de Henares: acercamiento a sus usos gráficos. En R. Díaz Moreno y B. Almeida (Eds.), *Estudios sobre la historia de los usos gráficos en español* (pp. 167-210). Axac.

Almeida Cabrejas, B. (2017). Mujeres que escriben y sistema(s) gráfico(s) en el siglo XVI: acercamiento desde un corpus de cartas. En B. Almeida, R. Díaz Moreno y C. Fernández López (Eds.), *Cansada tendré a Vuestra Excelencia con tan larga carta». Estudios sobre aprendizaje y práctica de la escritura por mujeres en el ámbito hispánico (1500-1900)* (pp. 9-53). Axac.

Archer, D., Kytö, M., Baron, A. y Rayson, P. (2015). Guidelines for normalizing Early Modern English corpora: Decisions and justifications. *ICAME Journal*, 39(1), 5-24. <https://doi.org/10.1515/icame-2015-0001>

Blas Arroyo, J. L. (2012). Tras las huellas de la variación y el cambio lingüístico a través de textos de inmediatez comunicativa. Fundamentos de un proyecto de sociolingüística histórica. En J. M. García Martín (Dir.), *Actas del IX Congreso Internacional de Historia de la Lengua Española* (pp. 1737-1756). Iberoamericana Vervuert.

Burnard, L. (2014). *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*. OpenEdition Press. <https://doi.org/10.4000/books.oep.426>

Calderón Campos, M. (2019). La edición de corpus históricos en la plataforma TEITOK. El caso de 'Oralia diacrónica del español'. *Chimera: Romance Corpora and Linguistic Studies*, (6), 21-36.

Calderón Campos, M. y García-Godoy, M. T. (Dirs.). (2019-). *Oralia diacrónica del español (ODE)*. <http://corpora.ugr.es/ode>

Calderón Campos, M. y Vaamonde, G. (2020). Oralia diacrónica del español: Un nuevo corpus de la Edad Moderna. *Scriptum Digital*, (9), 167-189. <https://doi.org/10.5565/rev/scriptum.108>

Cano Aguilar, R. (1996). Lenguaje 'espontáneo' y retórica epistolar en cartas de emigrantes españoles a Indias. En T. Kotschi, W. Oesterreicher y K. Zimmermann (Eds.), *El español hablado y la cultura oral en España e Hispanoamérica* (pp. 375-404). Iberoamericana Vervuert.

Corpus Mallorca. (s. f.). *Corpus Mallorca. Documentos castellanos en archivos de las Islas Baleares*. <https://www.corpusmallorca.es/>

Culpeper, J. y Kytö, M. (2010). *Early Modern English dialogues. Spoken interaction as writing*. Cambridge University Press.

Davies, M. (Dir.). (2001). *Corpus del Español (Género/Histórico) (CdEhist)*. [www.corpusdelespanol.org/hist-gen/](http://www.corpusdelespanol.org/hist-gen/)

Davies, M. (2009). Creating useful historical corpora: A comparison of CORDE, the Corpus del español, and the Corpus do português. En A. Enrique-Arias (Ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus* (pp. 137-166). Iberoamericana Vervuert. <https://doi.org/10.31819/9783865278685-009>

Dollinger, S. (2004). 'Philological computing' vs. 'philological outsourcing' and the compilation of historical corpora: a Late Modern English test case. *Vienna English Working Papers (VIEWS)*, 13, 3-23.

Enrique-Arias, A. (Dir.). (2008). *Biblia Medieval*. <http://www.bibliamedieval.es>

Fradejas Rueda, J. M. (2019). *Cuentapalabras. Estilometría y análisis de textos con R para Filólogos*. <http://www.aic.uva.es/cuentapalabras>

González Sopeña, I. (2023). Confusión de sibilantes y neutralización -r/-l en los inventarios de bienes andaluces del siglo XVIII. *Études romanes de Brno*, 44(1), 141-162. <https://doi.org/10.5817/ERB2023-1-9>

Grupo de investigación de textos para la historia del español (GITHE). (s. f.). *CODEA+* 2022. <https://www.corpuscodea.es/>

Hernangómez, D. (2024). *mapSpain: Administrative Boundaries of Spain*. <https://doi.org/10.5281/zenodo.5366622>

Honkapohja, A., Kaislaniemi, S. y Marttila, V. (2009). Digital Editions for Corpus Linguistics: Representing Manuscript Reality in Electronic Corpora. En A. H. Jucker, D. Schreier y M. Hundt (Eds.), *Corpora: Pragmatics and Discourse* (pp. 451-475). Rodopi.

Janssen, M. (2016). TEITOK: Text-Faithful Annotated Corpora. En N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk y S. Piperidis (Eds.), *Proceedings of the LREC* (pp. 4037-4043). European Language Resources Association.

Jenset, G. B. y McGillivray, B. (2017). *Quantitative Historical Linguistics. A Corpus Framework*. Oxford University Press.

Lass, R. (1997). *Historical linguistics and language change*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620928>

McEnery, T. y Wilson, A. (2001). *Corpus linguistics. An introduction*. Edinburgh University Press.

Meurman-Solin, A. (2001). Structured Text Corpora in the Study of Language Variation and Change. *Literary and Linguistic Computing*, 16(1), 5-27. <https://doi.org/10.1093/lrc/16.1.5>

Morala Rodríguez, J. R. (Dir.). (2018a). *Corpus léxico de inventarios (CorLexIn)*. <https://corlexin.unileon.es/>

Morala Rodríguez, J. R. (2018b). El proyecto CorLexIn y la variación diatópica en el léxico del Siglo de Oro. En D. Corbella Díaz, A. Fajardo Aguirre y J. Langenbacher-Liebgott (Eds.), *Historia del léxico español y Humanidades digitales* (pp. 397-417). Peter Lang.

Morala, J. R. y Perdiguer, H. (2019). La isoglosa de la aspiración de /f/ en el siglo XVII. En M. Castillo Lluch y E. Diez del Corral (Eds.), *Reescribiendo la historia de la lengua española a partir de la edición de documentos* (pp. 175-199). Peter Lang.

Núñez-Méndez, E. (2021). An overview of the sibilant merger and its development in Spanish. En E. Núñez-Méndez (Ed.), *Sociolinguistic approaches to sibilant variation in Spanish* (pp. 9-72). Routledge. <https://doi.org/10.4324/9781003153948-3>

Oesterreicher, W. (2004). Textos entre inmediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro. En R. Cano Aguilar (Ed.), *Historia de la lengua española* (pp. 729-769). Ariel.

Ortiz Cruz, D. (2020). Variaciones fonéticas en el léxico de inventarios de bienes aragoneses del siglo XVIII. *RILEX. Revista sobre investigaciones léxicas*, 3(1), 74-95. <https://doi.org/10.17561/rilex.3.1.5297>

P. S. Post Scriptum. (2014). *P. S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna*. <http://ps.clul.ul.pt>

Real Academia Española. (s. f.). *Corpus Diacrónico del Español (CORDE)*. <http://www.rae.es>

Real Academia Española. (2013). *Corpus del diccionario histórico de la lengua española (CDH)*. <http://www.rae.es>

Rodríguez Puente, P. (2018). En busca de lo hablado en lo escrito en los corpus diacrónicos del español: Una comparativa con los corpus anglosajones. *E-Scripta Romanica*, 5, 89-127. <https://doi.org/10.18778/2392-0718.05.09>

Rojo, G. (2010). Aguja de navegar corpus. En V. Castel y L. Cubo de Severino (Eds.), *La renovación de la palabra en el bicentenario de la Argentina. Los colores de la mirada lingüística* (pp. 1151-1163). Editorial FFyL, UNCuyo.

Rojo, G. (2016). Corpus textuales del español. En J. Gutiérrez-Rexach (Ed.), *Enciclopedia de Lingüística Hispánica* (Vol. 2, pp. 285-296). Routledge. <https://doi.org/10.4324/9781315713441-99>

Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge. <https://doi.org/10.4324/9781003119760>

Sánchez-Prieto Borja, P. (2020). Consonantismo avanzado en Madrid entre los siglos XVI y XIX a partir de un corpus documental (ALDICAM-CM). *Verba: Anuario gallego de filología*, 47, 55-86. <https://doi.org/10.15304/verba.47.5616>

Sánchez-Prieto Borja, P., Torrens Álvarez, M. J. y Vázquez Balonga, D. (Coords.). (2021). *La lengua de Madrid a lo largo del tiempo*. Editorial Universidad de Sevilla. <https://doi.org/10.12795/9788447231140>

Sánchez-Prieto Borja, P. y Vázquez Balonga, D. (2019). *La beneficencia madrileña. Lengua y discurso en los documentos de los siglos XVI a XIX*. Editorial Complutense.

Smith, J. (1997). *An Historical Study of English. Function, Form and Change*. Routledge.

Ueda, H. (2019). Las grafías bilabiales sonoras <u>, <v> y <b> del español en relación con el fonema /f/ y el paradigma sibilante. En M. Castillo Lluch y E. Diez del Corral (Eds.), *Reescribiendo la historia de la lengua española a partir de la edición de documentos* (pp. 141-174). Peter Lang.

Vaamonde, G. (2018). Escritura epistolar, edición digital y anotación de corpus. *Cuadernos del Instituto Historia de la Lengua*, (11), 139-164. <https://doi.org/10.58576/cilengua.vii1.53>

Vázquez Balonga, D. y Sánchez-Prieto, B. (2015). ¿Seseo en el centro peninsular? *Revista de Historia de la Lengua Española*, (10), 201-207. <https://doi.org/10.54166/rhle.2015.10.07>



eus

FACULTAD DE FILOLOGÍA  
UNIVERSIDAD DE SEVILLA