

## Los corpus digitales en el proyecto *DiacOralEs*<sup>1</sup>

Digital corpora in the *DiacOralEs* project

Santiago Del Rey Quesada  
Universidad de Sevilla (Sevilla, España)  
sdelrey@us.es  
<https://orcid.org/0000-0001-9871-4498>

Elena Carmona Yanes  
Universidad de Sevilla (Sevilla, España)  
elecarya@us.es  
<https://orcid.org/0000-0002-0427-2770>

Recibido el 21/10/2023, aceptado el 11/2/2024, publicado el 18/10/2024

*Creative Commons Attribution 4.0 International*  
© 2024 Santiago Del Rey Quesada, Elena Carmona Yanes

### Cómo citar este artículo

Del Rey Quesada, Santiago, Elena Carmona Yanes 2024. Los corpus digitales en el proyecto *DiacOralEs*. *Studia linguistica romanica* 2024.12, 86-106. <https://doi.org/10.25364/19.2024.12.5>.

### Resumen

El proyecto *DiacOralEs* estudiará entre 2022 y 2026 diversos aspectos relacionados con la caracterización discursivo-tradicional del español en distintas épocas, con especial atención a la influencia de otras lenguas mediante la traducción. Con este objetivo, se desarrollarán corpus digitales como *CorDisDial* (diálogos) y *EHA-PRESTUS* (prensa). Estos corpus se diseñarán en coherencia con un planteamiento que pretende superar algunas limitaciones de la lingüística de corpus para explotar el material textual empírico de forma que se armonicen el aprovechamiento de las herramientas tecnológicas actuales y las ventajas de un método filológico más tradicional que sigue siendo válido para interpretar el comportamiento de los hablantes y para analizar los factores potenciales que influyen sobre él.

### Abstract

The *DiacOralEs* project, running from 2022 to 2026, examines various aspects of the discursive-traditional characterization of Spanish across different periods, with special emphasis on the influence of other languages through translation. To this end, digital corpora such as *CorDisDial* (dialogues) and *EHA-PRESTUS* (press) are developed. These corpora are designed according to an approach that aims to address some limitations of corpus linguistics, enabling the exploitation of empirical text material in a way that integrates modern technological tools with the strengths of traditional philological methods, which still remain essential for interpreting speaker behavior and analyzing potential influencing factors.

---

<sup>1</sup> Esta publicación es parte del proyecto de I+D+i PID2021-123763NA-I00 *Hacia una diacronía de la oralidad/escrituralidad: variación concepcional, traducción y tradicionalidad discursiva en el español y otras lenguas románicas (DiacOralEs)*, financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

**Índice**

1 El proyecto <i>DiacOralEs</i> .....	88
2 Lingüística de corpus y traducción: los corpus comparables.....	89
2.1 Los <i>corpus-based translation studies</i> y la lingüística de corpus.....	89
2.2 Definición, directrices metodológicas y objetivo de los corpus comparables.....	90
2.3 El aprovechamiento de los corpus paralelos y comparables para la lingüística histórica románica.....	92
2.4 Corpus, intuición e ilustratividad.....	93
3 Dos corpus en <i>DiacOralEs</i> .....	96
3.1 El corpus <i>CorDisDial</i> como corpus comparable.....	96
3.2 Corpus <i>EHA-PRESTUS</i> : los documentos de la prensa anterior a 1850 y el influjo francófono.....	97
4 Consideraciones finales.....	101
Abreviaturas y referencias bibliográficas.....	102

## 1 El proyecto *DiacOralEs*

[1] En la estela de proyectos desarrollados por nuestro grupo de investigación entre 2015 y 2022 liderados por Araceli López Serena y que se centraban en la historia del español moderno<sup>2</sup>, entendida esta como una etapa fundamental para la evolución, el desarrollo e incluso la creación de numerosas estrategias discursivo-tradicionales que caracterizan diferentes géneros, entre ellos el periodístico, el proyecto *DiacOralEs*, financiado por el Ministerio de Ciencia e Innovación y la Agencia Estatal de Investigación, echó a andar precisamente a finales de 2022 y se prolongará hasta 2026. El nuevo proyecto recubre un espectro cronológico más amplio, pero sigue interesándose por la caracterización discursivo-tradicional del español en diferentes estadios lingüísticos y por la influencia de otras lenguas en la configuración discursiva de los géneros escritos a partir de la traducción.

[2] De entre las premisas teórico-metodológicas que sustentan el proyecto, conviene destacar, en primer lugar, el concepto de *variación*, que se entiende como elemento definidor de la historicidad lingüística (cf. Coseriu 1988 [1958]). Por lo que respecta a la descripción de los fenómenos estudiados en diferentes testimonios de la historia de las lenguas románicas, el proyecto se interesa especialmente por aspectos pragmático-discursivos en textos pertenecientes a diferentes tipologías textuales, incluso en períodos infrarrepresentados en la investigación diacrónica de la lengua, como es el caso del español moderno.

[3] Una novedad del proyecto *DiacOralEs* respecto de los proyectos precedentes de nuestro grupo de investigación es el papel protagonista que se concede a la traducción como fuente del conocimiento lingüístico y como responsable de la difusión de elementos discursivo-tradicionales<sup>3</sup>. Frente a lo que suele ser más habitual en la investigación histórica sobre textos traducidos, no asumimos que el calco desde el texto fuente (TF) sea el único fenómeno relevante, digno de análisis, en los textos meta (TM), sino que partimos de una propuesta, basada en ideas previas de Weinreich (1953), Coseriu (1977) y Kabatek (2000 [1996], 1997a, 1997b), centrada en el concepto de interferencia. Partimos del postulado metodológico de que toda unidad que pueda considerarse variante y no invariante en un TM es producto de la interferencia, que puede ser positiva (cuando provoca resultados marcados, bien perceptibles, en los textos) y negativa (cuando resulta en soluciones no marcadas, que seguramente pasarían desapercibidas si no se pudiera

---

2 *Tradiciones discursivas, tradiciones idiomáticas y unidades de análisis del discurso en la historia del español moderno* (FFI2014-51826-P), desarrollado entre 2015 y 2018, y *Tradicionalidad discursiva e idiomática, sintaxis del discurso, traducción y cambio lingüístico en la historia del español moderno: prosa (pre-)periodística/ensayística y literaria* (PGC2018-097823-B-I00), iniciado en 2019 y completado en 2022.

3 Sobre el concepto de tradicionalidad discursiva, cf. Koch (1987), Kabatek (2015), Winter-Froemel et al. (2015), Octavio de Toledo y Huerta (2018), López Serena (2021) y Del Rey Quesada (2021), entre otros.

confrontar el TM con el TF)<sup>4</sup>. Para determinar el peso que tienen ambos tipos de interferencia en la configuración discursiva del español a lo largo de la historia, el proyecto *DiacOralEs* se marca como objetivo la creación de dos corpus: uno de literatura dialógica de los siglos XV-XVII y otro de prensa de los siglos XVII-XIX, a la vez paralelos (que incorporan TF latinos o romance y los TM romances correspondientes) y comparables (que yuxtaponen a los TM traducidos textos no traducidos) (cf. § 3).

[4] La relación entre la lingüística de corpus y la traducción es intensa desde hace décadas y ha redundado en un gran avance en el conocimiento de los procedimientos traductivos que llevan a la conformación de los TM en diferentes géneros y tipos textuales, lo que nos ha permitido tener una idea más completa de los entresijos del contacto lingüístico en el ámbito de la escritura (cf. Baker 1995; Del Rey Quesada, Del Barrio de la Rosa & González Gómez 2018). En los siguientes apartados reflexionaremos sobre estas cuestiones y nos centraremos en la definición y significación de los corpus paralelos y comparables como los que el proyecto *DiacOralEs* se propone.

## 2 Lingüística de corpus y traducción: los corpus comparables

### 2.1 Los *corpus-based translation studies* y la lingüística de corpus

[5] Dentro de la lingüística de corpus, una corriente que ha gozado de especial adhesión en los últimos tiempos, sobre todo en el ámbito de los estudios de traducción e interpretación, es la de los denominados *corpus-based translation studies* (CTS). Baker (1995, 1996, 1998), abanderada de esta corriente, parte de la teoría polisistémica, de raigambre literaria, formulada por Even-Zohar (1978) al final de los años setenta, así como de los *descriptive translation studies*, para hacer suyos los corolarios de estos paradigmas, a saber, que las traducciones son hechos que pertenecen a la cultura meta, que pueden incluso constituir (sub)sistemas propios y que las traducciones son textos de pleno derecho, no solo representaciones o manifestaciones de otros textos (cf. Del Rey Quesada 2015).

[6] Partiendo de estas premisas, los estudiosos adeptos a los CTS se preocupan por la recopilación de textos literarios y no literarios contemporáneos para su explotación en corpus, corpus que poseen características específicas que los diferencian de los corpus monolingües al uso. Baker (1995: 233-234) distingue tres tipos de corpus que habría que tener en cuenta en el marco de los CTS:

- Corpus paralelos: textos de la lengua fuente en lengua A, y sus versiones traducidas en lengua B.
- Corpus multilingües: conjunto de corpus monolingües en distintas lenguas elaborado según criterios de diseño semejantes.

---

4 Para una exposición más pausada de estos tipos de interferencia, ver Del Rey Quesada (2016a, 2018a, 2020).

- Corpus comparables: dos colecciones separadas de textos en la misma lengua: un corpus de textos originales en la lengua en cuestión y otro que integra traducciones en esa lengua desde una(s) lengua(s) dada(s).

Los corpus paralelos y los corpus multilingües suponen herramientas muy interesantes para la lingüística románica, pues el contexto de comparabilidad que ofrecen los convierten en laboratorios de análisis idóneos para estudiar la variación lingüística en general, y, particularmente en el caso de los corpus paralelos, la variación diacrónica, como ha puesto de manifiesto un corpus bien conocido entre los historiadores de la lengua española como el corpus *Biblia medieval*, dirigido por Andrés Enrique-Arias (2009a, 2009b). Los corpus paralelos permiten al investigador seguir la pista del cambio lingüístico en un contexto de variación idéntico en diferentes épocas; de ahí su rentabilidad para el estudio histórico del español y de otras lenguas romances (cf. Del Rey Quesada 2016b).

## 2.2 Definición, directrices metodológicas y objetivo de los corpus comparables

[7] En cuanto a los corpus comparables (CC), según Baker (1995: 233), estos «effect a shift in the focus of theoretical research in the discipline, a shift away from comparing either ST [source text] with TT [target text] or language A with language B to comparing text production per se with translation». Además, Baker (1995) advierte de que estos corpus deben cubrir un dominio, variedad de lengua y período similares, así como de que los dos subcorpus (el de los textos traducidos y el de los textos no traducidos) deben poseer una longitud parecida y de que estos deben ser representativos en lo que se refiere al rango de los autores de textos originales y a los traductores. Para Baker (1995), la principal contribución que hace este tipo de corpus es la de identificar pautas específicas de los textos traducidos, independientemente de cuál sea la lengua de origen de cada texto. El interés, pues, reside en el proceso de traducción en sí, como actividad cognitiva que se refleja en la selección de determinadas estrategias compartidas entre distintos sistemas lingüísticos.

[8] Los representantes de los CTS se preocupan especialmente por homogeneizar los criterios de comparabilidad de los CC. Laviosa (1997) propone diferentes niveles en la fase de planificación del corpus, en los que habría que tener en cuenta diversos modos de producción y de presentación de los textos. En este trabajo, la autora se decanta por lo que considera la opción más idónea para el diseño de un CC del inglés.

[9] En el primer nivel, es importante clarificar si optamos por un corpus compuesto por textos íntegros, por extractos de textos, por una mezcla de estos dos tipos o por textos monitorizados (es decir, que se van actualizando periódicamente con ayuda de servidores diseñados específicamente para esta función). Asimismo, tendríamos que decidir si queremos un corpus diacrónico o sincrónico; general o terminológico; monolingüe, bilingüe o multilingüe; qué lengua o qué lenguas deseamos incorporar, y si el corpus va a incluir textos escritos u orales o am-

bos (hay que tener en cuenta que con esta distinción Laviosa (1997) se refiere a la dicotomía puramente medial, entre código gráfico y código fónico, y no a la concepcional en el sentido de Koch & Oesterreicher 2007 [1990]).

[10] El nivel 2, que ya depende de ciertas elecciones operadas en el nivel 1, se refiere a la distinción entre corpus sencillo y comparable en el caso de los corpus monolingües. En el nivel 3 de planificación del corpus, se establece una diferencia en función de si la elección lleva a un corpus sencillo, en cuyo caso tendríamos que distinguir entre un corpus hecho de textos traducidos o no traducidos, o lleva a un corpus comparable. En este caso, el que más nos interesa aquí, Laviosa propone una decisión metodológica relevante: la de si el CC es dependiente de los textos traducidos, es dependiente de los textos no traducidos o es independiente. Esta decisión condiciona la selección de textos y tipos textuales.

[11] Los parámetros del nivel 4 afectan a cuestiones relativas a los textos traducidos de los CC: cuántas lenguas fuente están implicadas; qué modos traductivos se encuentran (esto se refiere a si los textos se traducen de una sola lengua o de más); qué métodos traductivos aparecen (en este caso, se trata de distinguir si el texto traducido escrito se traduce de un texto escrito, si se traduce de una transcripción oral, si el texto traducido oral se traduce de un texto escrito o si se combinan varios de estos métodos en el corpus); cuál es el estatuto de la lengua meta respecto del traductor (si es lengua materna, lengua aprendida, etc.); cuál es el grado de especialización del traductor, y, por último, qué naturaleza tienen los textos incluidos (publicados o no). Como puede comprobarse, y aunque la propia Laviosa afirma que la lista de parámetros no pretende ser exhaustiva, mediante ellos se procura conseguir un corpus lo más homogéneo posible.

[12] El siguiente paso en la confección del corpus sería la delimitación de la categoría textual del texto. Conseguir la homogeneidad en este parámetro es sumamente complicado, como la propia autora confiesa (Laviosa 1997: 314), pues, aunque en el CC que ella propone incluye prosa literaria y prensa, reconoce que en este segundo tipo es más fácil conseguir textos parecidos en su forma y contenido que los textos de creación, por mucho que pertenezcan al mismo género.

[13] Es necesario precisar que el diseño y la confección de corpus comparables, tal y como los conciben los representantes de los CTS, están supeditados a unas premisas teóricas indisociables de los objetivos que justifican la creación de dichos corpus. Así, Laviosa (1997: 290) argumenta que el objetivo de los CC es hallar rasgos que permitan reconocer el estatus traducido versus no traducido de los textos:

One of the key features of this type of corpus is the comparability between its translational and non-translational components, which should be similar in as many respects as possible in an attempt to ensure that any linguistic differences found between them can be reliably attributed to their different status as translation vs. non-translation, rather than to confounding variables.

Por su parte, Baker (1993: 243) ya había insistido en que los CTS se interesan, sobre todo, por identificar características universales de la traducción<sup>5</sup>:

The most important task that awaits the application of corpus techniques in translation studies, it seems to me, is the elucidation of the nature of translated text as a mediated communicative event. In order to do this, it will be necessary to develop tools that will enable us to identify universal features of translation, that is features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems.

Así pues, los CC por los que abogan los investigadores pertenecientes a la corriente de los CTS se conciben como herramientas para el rastreo de regularidades y tendencias universales que permitan caracterizar el proceso de traducción *per se*, más allá de las lenguas de origen y de destino implicadas. Desde el punto de vista de la lingüística histórica románica, este propósito encuentra el escollo de reconocer a cada lengua características idiomáticas y discursivo-tradicionales específicas que impiden el establecimiento de universales aplicables a los textos traducidos como productos históricos definidos por la relación de contacto lingüístico entre lenguas que presentan frecuentemente un diferente estatus cultural (piénsese, por ejemplo, en las diferencias traductivas que se puede establecer entre las traducciones verticales (por ejemplo, latín > romance) y horizontales (romance > romance), según la terminología de Folea (1991 [1973]), en la Edad Media y el Renacimiento). Con todo, creemos que, aunque el objetivo para su creación no sea compartido por los CTS, la elaboración de CC también puede ser muy positiva para la lingüística histórica románica.

### 2.3 El aprovechamiento de los corpus paralelos y comparables para la lingüística histórica románica

[14] La creación de corpus comparables que integren textos traducidos y textos no traducidos de épocas pasadas puede resultar muy beneficioso para la historia de las lenguas romances y, muy particularmente, para la historia de la traducción. Tal y como pretendemos en el proyecto *DiacOralEs*, sería aún más atractivo y rentable para el análisis histórico la combinación, en un mismo corpus vertebrado en torno a un tipo textual específico o en torno a una época concreta, de corpus paralelos y corpus comparables.

[15] Un corpus de este tipo permitiría confrontar formas en relación con el análisis de la primera historicidad (gramática, vocabulario de una lengua particular) y de la segunda – en terminología de Kabatek (2015) –, esto es, las que son determinadas por las tradiciones discursivas que se actualizan en los textos. No se nos escapa, sin embargo, que el diseño de un corpus diacrónico de estas características es complejo: tendríamos que decidir qué textos caben en diferentes géneros, literarios o no y si la extensión de los textos originales y traducidos incluidos

---

<sup>5</sup> Para una discusión crítica de los diferentes universales traductológicos que se han propuesto en el ámbito de los CTS, ver Del Rey Quesada (2015).

puede ser variable. Debería, además, incorporar múltiples versiones de un mismo TF, tanto para la investigación sincrónica como para la diacrónica, y no siempre es fácil encontrar un TF que satisfaga este requisito – ya se ha demostrado la viabilidad del análisis, en este sentido, de las traducciones bíblicas (Enrique-Arias 2009b), de las traducciones ovidianas (Del Rey Quesada 2016c), de la *Bula ineffabilis* (cf. *Corpus Ineffabilis Deus*), pero no tenemos aún muchos más ejemplos explotados. Asimismo, con el objetivo de poder falsar las hipótesis de los CTS, un corpus comparable y a la vez paralelo también debería incluir textos traducidos y no traducidos escritos por el mismo autor y, para no errar en la adscripción idiomática o discursivo-tradicional de un fenómeno, debería contemplar varias lenguas fuente y varias lenguas meta.

[16] Resulta evidente que muchas de estas características entran en contradicción con la definición propia de los CC, que según Baker (1995) tienen que ser, como se apuntó en § 2.2, homogéneos en cuanto al período histórico considerado y a la extensión de los testimonios incluidos. Con todo, la idea de los CC es aprovechable en el ámbito de la lingüística histórica románica, aun cuando la magnitud del empeño orientado a crear corpus de las características que hemos enumerado en el párrafo precedente supone un reto de investigación considerable. Por ello, quizás sería conveniente empezar con CC diacrónicos que cumplan solo algunas de ellas. El objetivo, en cualquier caso, es que un CC diacrónico como el que concebimos en estas páginas pueda responder al análisis de los fenómenos de interferencia entre convergencia y divergencia en los textos traducidos; que permita rastrear elementos de tradicionalidad discursiva que parten de los TF o que nacen en los TM con independencia de aquellos; que habilite un análisis idiolectal o de estilo, por más que este hecho contravenga los principios teórico-metodológicos de los CTS, y que incorpore un aparato de metadatos lo suficientemente extenso como para analizar las tendencias traductivas en el contexto de escritura de cada texto. Los corpus que se han proyectado en *DiacOralEs* intentarán responder a estas necesidades del estudio histórico de los textos traducidos.

#### 2.4 Corpus, intuición e ilustratividad

[17] Sin pasar por alto las consideraciones metodológicas para la confección de corpus comparables expuestas más arriba, ni tampoco los presupuestos teóricos generales de la corriente de la lingüística de corpus, un rasgo común al planteamiento que sustenta el trabajo con los textos en el proyecto *DiacOralEs* es la idea de que la tarea de crear un corpus digital no tiene por qué estar primordialmente destinada al tratamiento de *big data* para análisis puramente estadísticos de los fenómenos. Los recursos digitales pueden ponerse al servicio de una metodología filológica más tradicional que no está necesariamente desfasada, sino que, por el contrario, sigue siendo muy necesaria para la obtención de conclusiones rigurosas sobre los comportamientos lingüísticos. Como explica López Serena (2021: 9), en el campo de la lingüística histórica se constata un rechazo predominante hacia la



conversión de la disciplina «en una lingüística de corpus que no tenga en cuenta los factores discursivo-tradicionales que condicionan la interpretación de cualquier dato lingüístico». En primer lugar, para trabajar desde la óptica de la tradicionalidad discursiva, es necesario que los corpus se puedan consultar leyendo la integridad de los textos, independientemente de que los acompañe cualquier elemento de etiquetado.

[18] El marco teórico de la lingüística de las variedades refrenda, además, la validez de los métodos de análisis cualitativos para el estudio de los textos, sobre la base, entre otras, de un concepto específico de intuición<sup>6</sup> como herramienta de aproximación científica a determinadas materias de estudio. De hecho, en opinión de Kabatek (2014: 713), incluso cuando el método incorpora algún componente cuantitativo, las intuiciones deben preceder a los cálculos: «después se harán estudios cuantitativos, pero se harán desde la intuición de lo que puede ser probable y lo que no». Esta priorización de lo cualitativo<sup>7</sup> debe ser puesta en relación con el problema de la representatividad potencial atribuida a los corpus lingüísticos. Para Kabatek (2013: 1), «un corpus representativo para la historia de la lengua es una construcción teóricamente imposible ya que la lengua, aunque solo se manifieste en textos, no es la suma de los textos, sino algo distinto». Así, no existiría «representatividad cuantitativa de un corpus, lo que (teóricamente) podría existir es una cierta representatividad cualitativa: un corpus con (casi) toda la gama de posibilidades de una lengua en un momento dado» (Kabatek 2013: 9).

[19] En este sentido, frente a la posibilidad de clasificar los estudios *con* corpus textuales que no se atienen estrictamente a los estándares del cálculo estadístico dentro de un cajón secundario o manifiestamente apartado de la ortodoxia de una lingüística *de* corpus basada en la representatividad estadística, Araceli López Serena se ha mostrado partidaria recientemente de recurrir a la noción de ilustratividad (López Serena & Del Rey Quesada 2022). Se trataría de una propiedad

---

6 Respecto al empleo técnico del vocablo *intuición* en epistemología lingüística y sus implicaciones para el modo de proceder en esta disciplina, pueden consultarse las contribuciones al número monográfico de la revista *Rilce. Revista de filología hispánica* dirigido por Araceli López Serena, quien, siguiendo a Itkonen (1976, 2008 [2003]), lo define como «el acto epistémico que da acceso al conocimiento de convenciones sociales como las normas lingüísticas» y que es característico del conocimiento de agente o conocimiento desde el interior «propio de las ciencias humanas, por oposición al conocimiento de observador que se ejerce en las ciencias naturales» (López Serena 2014: 692). Cf. ahora también López Serena (2019).

7 No sostenemos que la aproximación cualitativa, en cuanto etapa del procedimiento metodológico, haya de ser necesariamente la primera desde el punto de vista cronológico. Son distintas las vías exploratorias, cualitativas y cuantitativas que permitirán emitir hipótesis de partida. Sí defendemos la prioridad del componente intuitivo o cualitativo en la medida en la que no es prescindible como cauce interpretativo en este modelo: para la interpretación del comportamiento de los fenómenos variacionales, el acceso a la integridad de los textos y la recontextualización histórica de las situaciones comunicativas en las que se generaron no serán sustituibles por los meros resultados de los recuentos que puedan acompañar a estas interpretaciones.

o un criterio de composición para recopilaciones textuales que, más allá de las limitaciones materiales para la obtención de grandes extensiones de texto que puedan imponer una determinada época o ámbito documental<sup>8</sup>, se han proyectado para abordar objetos de estudio como la tradicionalidad discursiva y otros aspectos que resultan del comportamiento agentivo de los hablantes que, por su naturaleza, difícilmente podrían ser explicados a través de las predicciones cuantitativas de la estadística.

[20] El aprovechamiento de las ventajas del soporte digital y de algunas funcionalidades de recuento puede ponerse, por tanto, al servicio de un tipo de cuantificación descriptiva susceptible de aportar una información importante para la interpretación de algunos datos. Como ha indicado Del Rey Quesada (2021: 404), para trabajar desde la óptica de la tradicionalidad discursiva hay que tener en cuenta que «los elementos discursivo-tradicionales [caracterizadores] pueden ser tanto cualitativos como cuantitativos». Así, habría elementos que no son exclusivos de un tipo de texto, pero que «cuantitativamente funcionan como marcas discursivo-tradicionales de este tipo de textos» (Del Rey Quesada 2021: 404). Se entiende, de este modo, la utilidad de proyectar corpus digitales codificados, con anotación morfosintáctica y discursiva, que permitan la recuperación selectiva de información. Por ejemplo, el marcado de determinadas operaciones de traducción cuya presencia es objetiva no implica ningún tipo de interpretación de la intencionalidad del traductor. Es el caso, principalmente, de las adiciones y las supresiones. Aunque la simple cifra de la cantidad en la que se produzcan no permite extraer, *per se*, conclusiones definitivas, su localización previa puede agilizar distintos tipos de estudios cualitativos.

[21] En definitiva, defendemos que es necesario desarrollar recopilaciones textuales y corpus digitales que partan de un conocimiento profundo del medio discursivo que pretendan estudiar y de sus circunstancias condicionantes, con el objetivo, entre otros, de hacer avanzar la reflexión sobre cómo los corpus digitales pueden abordar no solo la diversidad de tipos de texto que vehicula cualquier lengua, sino también el carácter multicomposicional de los textos, que a menudo resulta determinante para explicar la distribución de los usos lingüísticos.

---

<sup>8</sup> Como es bien sabido, en lingüística histórica el uso de *software* de reconocimiento óptico no siempre es una opción eficiente. Las numerosas labores de corrección que implicaría el uso de este tipo de sistemas hacen que a menudo sea preferible una transcripción manual de los documentos, al menos cuando no es un objetivo prioritario la creación de un corpus muy extenso. Además, para los intereses de la lingüística es indispensable que esta transcripción y sus subsiguientes ediciones se sometan a revisiones minuciosas y detenidas, lo que, obviamente, multiplica el tiempo de trabajo requerido para el tratamiento previo a la publicación de un determinado volumen de texto.

### 3 Dos corpus en *DiacOralEs*

#### 3.1 El corpus *CorDisDial* como corpus comparable

[22] El corpus *CorDisDial* se concibe como un repositorio de textos dialógicos desde el siglo XV al XIX. En la primera fase del proyecto, en la que estamos trabajando ahora, se están preparando para su inclusión en el corpus diferentes traducciones tardomedievales y renacentistas de los diálogos de Cicerón, algunas versiones áureas de las comedias de Plauto y Terencio, algunos diálogos de Séneca (siglos XV-XVII) y varias traducciones del siglo XVI de los *Colloquia* de Erasmo. Mediante la elaboración del corpus se pretende ofrecer a la comunidad científica un conjunto ilustrativo de textos que permita reconocer dinámicas discursivo-tradicionales típicas del discurso dialógico en diferentes épocas del castellano y de otras lenguas romances y que habilite un análisis traductológico de fácil acceso al investigador gracias a la contraposición de TF y TM.

[23] Metodológicamente, a la selección de los testimonios que forman parte del proyecto les sigue la transcripción de los textos. Condición indispensable para esto es la digitalización de algunos de los testimonios por parte de las correspondientes bibliotecas universitarias, estatales y locales que los custodian, la mayoría de ellas sitas en Europa. Por lo que respecta a las ediciones de los siglos XX y XXI de textos antiguos con cuyos autores se haya acordado licencia de reproducción de acuerdo con las imposiciones legales vigentes, se recurre al reconocimiento automático del texto por medio de la *optical character recognition*, que, en cualquier caso, requiere una revisión exhaustiva del texto por parte del investigador principal y los técnicos del proyecto. Este procedimiento puede ser utilizado, además, en el caso de impresos antiguos que presenten un buen estado de legibilidad. Asimismo, se prevé el uso de las herramientas ofrecidas por el equipo de *Transkribus*. Dichas herramientas deben contribuir a agilizar el proceso de transcripción, algo absolutamente vital cuando se proyecta realizar un corpus que aspire a la representatividad, si bien este no es un horizonte de referencia *sine qua non* en nuestro proyecto, que parte de la premisa de que la validez científica de los corpus puede descansar en la idea de ilustratividad más que en la de representatividad, habida cuenta de que, por mucho que un corpus incluya miles o millones de palabras, nunca será representativo de la totalidad de variedades de una lengua histórica.

[24] En el tratamiento del texto, los criterios de edición vienen determinados por las necesidades específicas del proyecto, que, en esta primera fase, procurará satisfacer las necesidades de alineación de los TF y TM correspondientes, sin que se descarte la anotación sintáctica y discursiva en futuras fases. Seguiremos, en este sentido, las recomendaciones del consorcio TEI y nos basaremos, más específicamente, en la tecnología *TEITOK*, que se ha revelado muy rentable para la ejecución de proyectos de filología románica como el que planteamos aquí (cf. la página *Projects* de *TEITOK*). Para criterios de edición más específicos – desarrollo

de abreviaturas, saltos de línea y de página, títulos de sección, etc. —, nos guiaremos por las propuestas de la red *CHARTA*, cuyos principios de edición digital se enmarcan en un esquema de uso abierto integrable con TEI.

[25] La segmentación del texto, necesaria para la creación del corpus paralelo anotado, se llevará a cabo atendiendo a las intervenciones de los personajes que participan en el diálogo, pero también a los enunciados de los que se compone cada intervención. Especialmente compleja resulta la segmentación de texto que no se reduce a una correspondencia exacta entre el TF y el TM en las distintas versiones que configuran el corpus paralelo, es decir, sobre todo cuando la configuración sintáctica de una de las versiones es muy diferente a la del texto original o a la de otras versiones, si bien la alineación está bien descrita en las guías de uso de TEI.

[26] Para la creación del corpus se utiliza el editor de textos en XML *Oxygen*, que nos sirve, con un alto grado de eficiencia, para la edición, la depuración, el etiquetado a todo nivel, el procesado y la interrogación del corpus. Además, *Oxygen* presenta la ventaja de conectarse nativamente con *eXistdb*, un programa que ya se ha revelado útil para el desarrollo de software específico para alineación de textos. Por lo demás, se prevé que la publicación del corpus también se realice dentro de las posibilidades de *eXistdb*, que nos permitirá desarrollar el software apropiado para generar una interfaz equilibrada que ofrezca al usuario la opción de contemplar y hacer búsquedas en las traducciones que desee de entre las que conforman *CorDisDial*.

### 3.2 Corpus *EHA-PRESTUS*: los documentos de la prensa anterior a 1850 y el influjo francófono

[27] Otro de los objetivos descriptivos principales de *DiacOralEs* es estudiar la primera etapa del desarrollo del discurso periodístico en España (ca. 1675-1850), un objetivo cuyo interés procede, por una parte, del hecho de que siguen siendo escasos, en general, los trabajos sobre el español peninsular posteriores al siglo XVII y, en particular, de la primera mitad del XVIII<sup>9</sup>; y, por otra parte, de la existencia de procesos de elaboración discursiva que pueden constatarse en la formación de este medio en el período que podríamos denominar protoperiodístico<sup>10</sup>.

9 En el ámbito del análisis gramatical y discursivo, disponemos de trabajos de sumo interés sobre distintas muestras textuales datadas entre 1680 y 1765, aproximadamente (cf., entre otros, Girón Alconchel 2002, 2003, 2004, 2008, 2012; Octavio de Toledo y Huerta 2007, 2016; Ramírez Luenigo 2013; Sáez Rivera 2003, 2004, 2008). Ninguno de ellos se centra en el ámbito periodístico, aunque sí lo tienen en consideración Octavio de Toledo y Huerta (2016), y Sáez Rivera & Octavio de Toledo y Huerta (2020).

10 A pesar de la importancia de este proceso, falta hoy en día el material empírico para estudiarlo en profundidad. Existe una evidente laguna documental para la prensa europea en *CORDE*, aunque, como es sabido, lo que está infrarrepresentado en este corpus no es tanto lo periodístico como el siglo XIX y, sobre todo, el XVIII. La consecuencia es que el recurso a *CORDE*, si nos da acceso a textos preperiodísticos, no permite una aproximación a la lengua que llamaríamos protoperiodís-

[28] Para esta época y para este proceso de elaboración, resulta fundamental un hecho fácilmente observable a nivel externo y superestructural, pero que ha recibido mucha menos atención desde un punto de vista microlingüístico: la acción constante e intensa de los modelos textuales extranjeros, muy especialmente francófonos, que se imitan, se reescriben o se traducen directamente como fuente habitual de contenidos y de documentos para las nuevas publicaciones periódicas.

[29] Así pues, con el fin de abordar la cuestión de una potencial influencia idiomática directa del francés sobre el español en este ámbito, que consideramos, como hipótesis inicial, posible pero no presumible<sup>11</sup>, hemos emprendido en el marco de *DiacOralEs* las labores de edición que permitirán componer un subcorpus específico de textos protoperiodísticos y de fuentes redactadas en francés: el corpus *EHA-PRESTUS*.

[30] *EHA-PRESTUS* albergará una muestra ilustrativa de la diversidad textual y de la evolución de los formatos periodísticos que tiene ya lugar entre finales del siglo XVII y mediados del XIX. Aunque será susceptible de ser ampliado en el futuro, dentro de los límites temporales de *DiacOralEs* podrán ser incluidos los siguientes testimonios:

1. La edición bilingüe (completa) de los números mensuales del *Mercurio Histórico y Político* en la época de Salvador Mañer (1738-1744), que se redacta íntegramente como la traducción del *Mercure historique et politique* publicado en esos momentos en La Haya por Jean Rousset. El *Mercurio* se convertirá, tras la *Gaceta*, en el segundo gran referente del periodismo de información política y bélica, y también en la primera cabecera que adquirirá carácter oficial al quedar incorporado a la Corona en 1756 (cf. Seoane & Saiz 2007: 35-36). Si ya en la *Gaceta* es posible rastrear algunos artículos traducidos de su precedente francés (Carmona Yanes 2019: 156), el *Mercurio* nace y se desarrolla durante sus primeras décadas de existencia como una ver-

---

tica, la de los formatos que ya se dan al público con una periodicidad regular. La hemeroteca digital de la Biblioteca nacional de España, a pesar de contar con algunas funcionalidades que permiten, por ejemplo, hacer búsquedas por palabra clave relativamente correctas, no puede considerarse un corpus propiamente dicho y queda, evidentemente, sin integrar con el resto de los tipos de textos representados en *CORDE* y otros corpus generales del español. La situación es diferente en *CORDIAM*, que se nutre a través de uno de sus tres subcorpus de abundantes textos periodísticos publicados en América. Como fondos anteriores a 1850, dispone al 29/9/2023 de 7663 textos (3141324 palabras), frente a los 379 textos (1288897 palabras) del *CORDE* para las mismas fechas. El desequilibrio es aún mayor si tenemos en cuenta que *CORDIAM* no incluye en esta categoría formatos preperiodísticos anteriores al siglo XVIII.

<sup>11</sup> Como Castillo Lluch (2005) o Del Rey Quesada (2018b) han constatado en otros estudios, en un texto traducido no puede darse por sentada, de manera apriorística, una tendencia hacia la convergencia con el original, sea por voluntad, sea por impericia del traductor, y encontramos, de hecho, traducciones que optan sistemáticamente por una estrategia divergente. Además, las coincidencias formales entre el TF y el TM no siempre pueden explicarse por una voluntad identificadora (Del Rey Quesada 2018b: 638).

- sión en español del original francófono (Carmona Yanes 2020).
2. Artículos del *Semanario Económico* de Pedro Saura (1765-1767). La prensa especializada en contenido técnico y científico es una de las señas de identidad del periodismo dieciochesco. En ella encontramos un exponente máximo de la hibridación discursiva típica del medio en esta época (Méndez Orense & Carmona Yanes 2023): la prensa literaria y erudita adaptará a sus formatos y a sus ritmos de publicación distintas formas textuales procedentes del ámbito académico, persiguiendo la fidelización de un público que, aunque minoritario, se procurará ampliar y diversificar. El caso del *Semanario* de Saura es un exponente de la primacía, a escala europea, de las fuentes francófonas como originales e intermediarias en la difusión de este tipo de saber utilitario (Carmona Yanes & Méndez Orense, en prensa).
  3. Artículos de las secciones de *Noticias extranjeras* de la prensa del segundo cuarto del siglo XIX. Tras las transformaciones que culminan en la década de 1830, estos periódicos se aproximarán ya más en su funcionamiento y en su aspecto a los actuales. El recurso a la traducción, aún muy frecuente, se concentrará en las secciones de información internacional. Las principales cabeceras de tirada nacional (*Diario del Comercio*, *El Eco del Comercio*, *El Español*, *El Nacional*, *El Castellano*, *El Correo Nacional*, *El Corresponsal*, *El Piloto* y *El Heraldo*, entre otras) a menudo aluden a las fuentes periódicas extranjeras de las que se nutren (*Journal des débats politiques et littéraires*, *Le Constitutionnel*, *Le Courrier français*, *L'Écho français*, *La Presse*, *Le Siècle*). Una antología de estos documentos se encuentra ya disponible en Carmona Yanes (2022), y quedará integrada en *EHA-PRESTUS*.

[31] Con esta selección de hitos del discurso protoperiodístico español traducido del francés confeccionaremos un corpus paralelo de textos fuente y textos meta pareados. Al margen de las posibles dificultades para la localización de algunos originales, estamos aquí ante casos en los que se practica un ejercicio de traducción propiamente dicha: una fuente extranjera es la base textual clara y fundamental para la redacción del texto español. Prescindimos, por el momento, de otros testimonios en los que pueden rastrearse labores más libres de reescritura.

[32] Para valorar la incidencia de los originales francófonos sobre las elecciones del traductor, se hace indispensable contar con corpus comparables, de textos producidos originalmente en español, a través de los que se pueda aislar el contacto interlingüístico como un factor potencial para las soluciones del redactor en el caso de los textos traducidos (cf. § 2.2). La tarea de establecer un corpus comparable para los textos protoperiodísticos españoles no está exenta de dificultades, pues estos textos se construyen en un ámbito donde los modelos documentales son siempre foráneos y donde se traduce con mucha frecuencia.

[33] En el caso del *Mercurio*, no existe en el español de su época un equivalente tipológico de estos artículos extensos de narración noticiosa de tema político y militar, ya que se trata de una publicación pionera en este sentido. Otras fórmu-

las periodísticas de la época presentan ya algunas diferencias compositivas importantes, pero, sobre todo, el problema de que, en la primera mitad del siglo XVIII, casi cualquier texto de la prensa es sospechoso de ser una traducción. Es posible establecer vínculos cercanos con el relato historiográfico, una vía de comparabilidad que no descartamos para posteriores estudios. Pero el hecho de que la obra provenga, como era habitual en la época, de un redactor único, nos permite recurrir al factor idiolectal como primer término de comparación. Salvador Mañera cuenta con una extensa producción textual originalmente concebida en español, previa a la etapa en la que comienza a dedicarse de manera regular a la traducción (Aguilar Piñal 1981-2001). Aunque menos prolífico, también es obra de un redactor único el *Semanario Económico* de Pedro Araus. En ambos casos y con las debidas precauciones podríamos acceder a las particularidades de sendos idiolectos<sup>12</sup> que se adaptan a la expresión prototípica del medio periodístico incipiente, y que pueden compararse ulteriormente con otros empleados en los géneros más próximos que llegaron a cultivarse en la época.

[34] En cuanto a la información internacional decimonónica, es mucho más sencillo componer un corpus comparable estrictamente periodístico. Aunque ya no es posible atender al factor idiolectal de manera tan evidente, se puede presuponer y comprobar que comparten redactores los textos de las mismas cabeceras. Se hace necesario, eso sí, utilizar textos del mismo género pero de una sección y temática diferentes: información nacional y no internacional, pues en esta última estaríamos de nuevo ante el riesgo de encontrar únicamente textos traducidos.

[35] No obstante, si el desarrollo de estos corpus comparables desempeña un papel fundamental para alcanzar los objetivos de investigación que perseguimos dentro del planteamiento de *DiacOralEs*, los textos originalmente redactados en español no se publicarán, al menos, en la primera etapa de existencia de *EHA-PRESTUS*, que se presentará como un corpus paralelo francés-español. Naturalmente, el interés que atribuimos a su publicación no es solo la transparencia respecto a las fuentes en las que se basan los análisis realizados por los investigadores de *DiacOralEs*. Más allá de la disponibilidad de documentos de prensa española anterior a 1850 que sigue requiriendo la lingüística diacrónica, consideramos especialmente la utilidad que en este material podrá encontrar un público meta de estudiosos de la historia del periodismo y de la historia de la traducción (donde precisamente la traducción de textos periodísticos constituye una laguna importante de los estudios diacrónicos), o, igualmente, en los ámbitos de la historia general y de otras disciplinas humanísticas. Para este tipo de investigadores, se marcarán las principales intervenciones materiales del traductor (omisiones, adiciones, transposiciones) y otras operaciones que pudieran dar una idea de cómo trabajaba el periodista en su rol de traductor. Avanzaremos también en las labores de lemati-

---

12 Lo cual representa, además, una ventaja para estudiar gran parte de los rasgos morfosintácticos abiertos a la variación que constituyen nuestro objeto de estudio principal (Barra Jover 2015).

zación que faciliten las búsquedas por palabra clave y el funcionamiento de las herramientas de recuento.

#### 4 Consideraciones finales

[36] La investigación diacrónica actual difícilmente puede prescindir de las herramientas digitales que proporcionan un acceso más fácil y sistemático a los fenómenos recurrentes que pueden observarse en distintos ámbitos textuales. La tarea de desarrollar los corpus que doten del material empírico necesario para objetivos de investigación diversos (y en concreto en el dominio de la variación gramatical y discursivo-tradicional) no está exenta, sin embargo, de dificultades técnicas pero también metodológicas y teóricas relacionadas con el tratamiento de los textos y de los resultados que cabe poder esperar de su análisis. Dentro de los planteamientos de *DiacOralEs*, los corpus paralelos *CorDisDial* y *EHA-PRES-TUS*, próximamente disponibles en acceso libre gratuito, se enfrentan en su diseño a los retos derivados del trabajo con documentos no contemporáneos que requieren pausadas labores de edición filológica y digital y un conocimiento profundo de la literatura dialógica y del medio periodístico, respectivamente, con vistas a instaurar un marco interpretativo para las elecciones concretas del traductor que permita explorar el influjo del texto fuente y su lengua como un factor variacional más.



### Abreviaturas y referencias bibliográficas

- Aguilar Piñal 1981-2001 = Francisco Aguilar Piñal 1981-2001. *Bibliografía de autores españoles del siglo XVIII*. Consejo Superior de Investigaciones Científicas.
- Baker 1993 = Mona Baker 1993. Corpus linguistics and translation studies – Implications and applications. Mona Baker, Gill Francis, Elena Tognini-Bonelli (eds.). *Text and technology. In Honour of John Sinclair*. Benjamins, 233-250.
- Baker 1995 = Mona Baker 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7.2, 223-243.
- Baker 1996 = Mona Baker 1996. Corpus-based translation studies: The challenges that lie ahead. Harold Somers (ed.). *Terminology, LSP and translation. Studies in language engineering in Honour of Juan C. Sager*. Benjamins, 175-186.
- Baker 1998 = Mona Baker 1998. Réexplorer la langue de la traduction: une approche par corpus. *Meta. Journal des traducteurs* 43.4, 480-485.
- Barra Jover 2015 = Mario Barra Jover 2015. Método y teoría del cambio lingüístico: argumentos en favor de un 'método idiolectal'. José M. García Martín, Teresa Bastardín Candón, Manuel Rivas Zancarrón (eds.). *Actas del IX Congreso Internacional de Historia de la Lengua Española (Cádiz, 2012). Vol. 1. Iberoamericana*, Vervuert, 263-292.
- Biblia medieval* = Andrés Enrique-Arias (ed.) 2008-. *Biblia medieval*. <http://bibliamedieval.es>.
- Carmona Yanes 2019 = Elena Carmona Yanes 2019. Los textos periodísticos traducidos del francés (1830-1845): Lengua y tradicionalidad. *Anuari de Filologia. Estudis de Lingüística* 9, 145-178. <https://doi.org/10.1344/AFEL2019.9.5>.
- Carmona Yanes 2020 = Elena Carmona Yanes 2020. Contactos entre el francés y el español en el discurso periodístico: la variación morfosintáctica en el *Mercurio Histórico y Político* en la época de Salvador Mañer (1738-1745). *Boletín Hispánico Helvético* 35-36, 87-121. <https://doi.org/10.36950/bhhd.vi35-36.10138>.
- Carmona Yanes 2022 = Elena Carmona Yanes 2022. *Fuentes francófonas en las secciones de Noticias extranjeras de la prensa española (ca. 1830-1845): una antología de textos traducidos, con sus originales*. Editorial Universidad de Sevilla. <https://dx.doi.org/10.12795/9788447224739>.
- Carmona Yanes & Méndez Orense, en prensa = Elena Carmona Yanes, María Méndez Orense, en prensa. El *Semanario Económico* en la época de Pedro Araus (1765-1767): difusión y traducción de fuentes francófonas. *Hermeneus*.
- Castillo Lluch 2005 = Mónica Castillo Lluch 2005. Translación y variación lingüística en Castilla (siglo XIII): la lengua de las traducciones. *Cahiers d'études hispaniques médiévales* 28, 131-144. [https://www.persee.fr/doc/cehm\\_0396-9045\\_2005\\_num\\_28\\_1\\_1697](https://www.persee.fr/doc/cehm_0396-9045_2005_num_28_1_1697).
- CC = Corpus comparables.
- CHARTA = Pedro Sánchez-Prieto Borja (ed.) 2011-. *Corpus hispánico y americano en la red*. <https://www.redcharta.es>.
- CORDE = Real Academia Española (ed.) 2008. *Corpus diacrónico del español*. <http://corpus-rae.es/cordenet.html>.
- CORDIAM = Concepción Company Company, Virginia Bertolotti (eds.) 2016-. *Corpus diacrónico y diatópico del español de América*. <https://www.cordiam.org/>.
- CorDisDial = Santiago Del Rey Quesada (ed.), en preparación. *Corpus del discurso dialógico en la historia de las lenguas romances*. Universidad de Sevilla.
- Corpus Ineffabilis Deus* = Johannes Kabatek, Manuela Crivelli (eds.) 2020-. *Corpus Ineffabilis Deus*. <https://corpus-ineffabilis.org>.
- Coseriu 1977 = Eugenio Coseriu 1977. Sprachliche Interferenz bei Hochgebildeten. Herbert Kolb, Hartmut Lauffer (eds.). *Sprachliche Interferenz. Festschrift für Werner Betz zum 65. Ge-*

- burtstag*. Niemeyer, 77-100.
- Coseriu 1988 [1958] = Eugenio Coseriu 1988 [1958]. *Sincronía, diacronía e historia. El problema del cambio lingüístico*. 3a edición. Gredos.
- CTS = *Corpus-based translation studies*.
- Del Rey Quesada 2015 = Santiago Del Rey Quesada 2015. Universales de la traducción e historia de la lengua: algunas reflexiones a propósito de las versiones castellanas de los *Colloquia* de Erasmo. *Iberoromania* 81.1, 83-102.
- Del Rey Quesada 2016a = Santiago Del Rey Quesada 2016a. Interferencia latín-romance en Alfonso X: la traducción como pretexto de la elaboración sintáctica. *La corónica. A journal of medieval Hispanic languages, literatures, and cultures* 44.2, 75-109.
- Del Rey Quesada 2016b = Santiago Del Rey Quesada 2016b. Traducción y tradición en los corpus: nuevas perspectivas para la lingüística histórica. Johannes Kabatek (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. De Gruyter, 40-56.
- Del Rey Quesada 2016c = Santiago Del Rey Quesada 2016c. Ocho tipos de lengua, cara a cara: las traducciones de la epístola ovidiana de Dido a Eneas en la Edad Media y el Siglo de Oro. Araceli López Serena, Antonio Narbona Jiménez, Santiago Del Rey Quesada (eds.). *El español a través del tiempo. Estudios ofrecidos a Rafael Cano Aguilar*. Vol. 1. Editorial Universidad de Sevilla, 415-439.
- Del Rey Quesada 2018a = Santiago Del Rey Quesada 2018a. El *De senectute* de Cicerón en romance (ss. XIV-XVI): un estudio sintáctico contrastivo. *Anuari de filologia. Estudis de lingüística* 8, 21-56. <https://doi.org/10.1344/AFEL2018.8.2>.
- Del Rey Quesada 2018b = Santiago Del Rey Quesada 2018b. Latinismo, antilatinismo, hiperlatinismo y heterolatinismo: la sintaxis de la prosa traducida erasmiana del Siglo de Oro. María L. Arnal Purroy et al. (eds.). *Actas del X Congreso internacional de historia de la lengua española. Zaragoza, 7-11 de septiembre de 2015*. Vol. 1. Institución Fernando el Católico, 623-645.
- Del Rey Quesada 2020 = Santiago Del Rey Quesada 2020. The analysis of linguistic variation in translation studies. A proposal for classifying translational phenomena between source text and target text. *Hikma. Revista de traducción* 19.1, 209-237. <https://doi.org/10.21071/hikma.v19i1.12333>.
- Del Rey Quesada 2021 = Santiago Del Rey Quesada 2021. *Grupos léxicos paratácticos en la Edad Media romance. Caracterización lingüística, influencia latinizante y tradicionalidad discursiva*. Peter Lang.
- Del Rey Quesada, Del Barrio de la Rosa & González Gómez 2018 = Santiago Del Rey Quesada, Florencio Del Barrio de la Rosa, Jaime González Gómez 2018. *Lenguas en contacto, ayer y hoy: traducción y variación desde una perspectiva filológica*. Introducción. Santiago Del Rey Quesada, Florencio Del Barrio de la Rosa, Jaime González Gómez (eds.). *Lenguas en contacto, ayer y hoy. Traducción y variación desde una perspectiva filológica*. Peter Lang, 9-24.
- DiacOralEs* = Santiago Del Rey Quesada (ed.) 2022-2026. *Hacia una diacronía de la oralidad/escrituralidad: variación concepcional, traducción y tradicionalidad discursiva en el español y otras lenguas románicas*. Proyecto I+D+i PID2021-123763NA-I00 financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE. Universidad de Sevilla.
- EHA-PRESTUS* = Elena Carmona Yanes (ed.), en preparación. *Corpus de textos periodísticos traducidos del grupo EHA* (Español hablado en Andalucía/Estudios históricos de análisis del discurso). Universidad de Sevilla.
- Enrique-Arias 2009a = Andrés Enrique-Arias 2009a. Introducción. Lingüística de corpus y diacronía de las lenguas iberorrománicas. Andrés Enrique-Arias (ed.). *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Iberoamericana, Ver-

- vuert, 11-21.
- Enrique-Arias 2009b = Andrés Enrique-Arias 2009b. Ventajas e inconvenientes del uso de *Biblia medieval* (un corpus paralelo y alineado de textos bíblicos) para la investigación en lingüística histórica del español. Andrés Enrique-Arias (ed.). *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Iberoamericana, Vervuert, 269-283.
- Even-Zohar 1978 = Itmar Even-Zohar 1978. The position of translated literature within the literary polysystem. James S. Holmes, José Lambert, Raymond van den Broeck (eds.). *Literature and translation. New perspectives in literary studies with a basic bibliography of books on translation studies*. Acco, 17-27.
- eXistdb = eXist Solutions 2013-. *eXistdb*. <http://exist-db.org>.
- Folena 1991 [1973] = Gianfranco Folena 1991 [1973]. *Volgarizzare e tradurre*. Einaudi.
- Girón Alconchel 2002 = José L. Girón Alconchel 2002. Procesos de gramaticalización del español clásico al moderno. María T. Echenique Elizondo, Juan P. Sánchez Méndez (eds.). *Actas del V Congreso internacional de historia de la lengua española. Valencia, 31 de enero-4 de febrero de 2000*. Vol. 1. Gredos, 103-122.
- Girón Alconchel 2003 = José L. Girón Alconchel 2003. Evolución de la cohesión en el discurso ensayístico entre 1648 y 1726. José L. Girón Alconchel et al. (eds.). *Estudios ofrecidos al profesor José Jesús de Bustos Tovar*. Vol. 1. Universidad Complutense, 331-360.
- Girón Alconchel 2004 = José L. Girón Alconchel 2004. *El hombre práctico* (1686) de Gutiérrez de los Ríos: el español de finales del XVII. Isaías Lerner, Robert Nival, Alejandro Alonso (eds.). *Actas del XIV Congreso de la Asociación internacional de hispanistas. New York, 16-21 de julio de 2001. 1. Literatura medieval, lingüística, historia, teoría literaria, estudios culturales*. Juan de la Cuesta, 251-264.
- Girón Alconchel 2008 = José L. Girón Alconchel 2008. La lengua de un embajador y un marino del siglo XVIII: ¿español moderno ya, o todavía clásico? Concepción Company Company, José G. Moreno de Alba (eds.). *Actas del VII Congreso internacional de historia de la lengua española. Mérida (Yucatán), 4-8 de septiembre de 2006*. Vol. 2. Arcos Libros, 2243-2253.
- Girón Alconchel 2012 = José L. Girón Alconchel 2012. El cambio y el no cambio gramatical en el relato histórico en la transición del siglo XVII al XVIII. *Cuadernos dieciochistas* 13, 29-49. <https://revistas.usal.es/dos/index.php/1576-7914/article/view/9511>.
- Itkonen 1976 = Esa Itkonen 1976. *Linguistics and empiricism. Answers to criticism*. University of Helsinki.
- Itkonen 2008 [2003] = Esa Itkonen 2008 [2003]. *¿Qué es el lenguaje? Introducción a la filosofía de la lingüística*. Biblioteca Nueva [traducción de *What is language. A study in the philosophy of linguistics*. Turun Yliopisto].
- Kabatek 1997a = Johannes Kabatek 1997a. Dime cómo hablas y te diré quién eres. Mezcla de lenguas y posicionamiento social. *Revista de antropología social* 6, 215-236. <https://revistas.ucm.es/index.php/RASO/article/view/RASO9797110215A>.
- Kabatek 1997b = Johannes Kabatek 1997b. Zur Typologie sprachlicher Interferenzen. Wolfgang W. Moelleken, Peter J. Weber (eds.). *Neue Forschungsarbeiten zur Kontaktlinguistik*. Dümmler, 232-241.
- Kabatek 2000 [1996] = Johannes Kabatek 2000 [1996]. *Os falantes como lingüistas. Tradición, innovación e interferencias no galego actual*. Xerais [traducción de *Die Sprecher als Linguisten. Interferenz- und Sprachwandelphänomene dargestellt am Galicischen der Gegenwart*. De Gruyter].
- Kabatek 2013 = Johannes Kabatek 2013. ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania* 77.1, 8-28.

- Kabatek 2014 = Johannes Kabatek 2014. Lingüística empática. *Rilce. Revista de filología hispánica* 30.3, 705-723. <https://doi.org/10.15581/008.30.366>.
- Kabatek 2015 = Johannes Kabatek 2015. Warum die 'zweite Historizität' eben doch die zweite ist – von der Bedeutung von Diskurstraditionen für die Sprachbetrachtung. Franz Lebsanft, Angela Schrott (eds.). *Diskurse, Texte, Traditionen. Modelle und Fachkulturen in der Diskussion*. V&R Unipress, 49-62.
- Koch & Oesterreicher 2007 [1990] = Peter Koch, Wulf Oesterreicher 1990 [2007]. *Lengua hablada en la Rumania. Español, francés, italiano*. Gredos [traducción de *Gesprochene Sprache in der Rumania: Französisch, Italienisch, Spanisch*. Niemeyer].
- Koch 1987 = Peter Koch 1987. *Distanz im Dictamen. Zur Schriftlichkeit und Pragmatik mittelalterlicher Brief- und Redemodelle in Italien*. Habilitationsschrift, Universität Freiburg.
- Laviosa 1997 = Sara Laviosa 1997. How comparable can 'comparable corpora' be? *Target* 9.2, 289-319
- López Serena & Del Rey Quesada 2022 = Araceli López Serena, Santiago Del Rey Quesada 2022. Oralidad y escrituralidad en los corpus paralelos de las lenguas romances: el proyecto *DiacOralEs*. *VI Congreso internacional de corpus diacrónicos en lenguas iberorrománicas. Venecia, 5-7 de octubre de 2022*. <https://venezia6codili.wordpress.com/libro-de-resumenes>.
- López Serena 2014 = Araceli López Serena 2014. Historia de la lengua e intuición. Presentación. *Rilce. Revista de filología hispánica* 30.3, 691-704. <https://doi.org/10.15581/008.30.365>.
- López Serena 2019 = Araceli López Serena 2019. *La lingüística como ciencia humana. Una incurción desde la filosofía de la ciencia*. Arcos Libros.
- López Serena 2021 = Araceli López Serena 2021. La tradicionalidad discursiva como materia y las tradiciones discursivas como objeto de estudio. *Verba. Anuario galego de filoloxía* 48, 1-40. <https://doi.org/10.15304/verba.48.6864>.
- Méndez Orense & Carmona Yanes 2023 = María Méndez Orense, Elena Carmona Yanes 2023. Tradicionalidad discursiva y variación morfosintáctica en la prensa económica de la Ilustración española. *Zeitschrift für romanische Philologie* 139.1, 88-123.
- Octavio de Toledo y Huerta 2007 = Álvaro S. Octavio de Toledo y Huerta 2007. Un rasgo sintáctico del primer español moderno (ca. 1675-1825): las relaciones interoracionales con *interin* (*que*). Marta Fernández, Araceli López (eds.). *Cuatrocientos años de la lengua del Quijote. Estudios de historiografía e historia de la lengua española. Actas del V Congreso nacional de la Asociación de jóvenes investigadores de historiografía e historia de la lengua española. Sevilla, 31 de marzo, 1 y 2 de abril de 2005*. Servicio de Publicaciones de la Universidad de Sevilla, 421-442.
- Octavio de Toledo y Huerta 2016 = Álvaro S. Octavio de Toledo y Huerta 2016. Antonio Muñoz y la sintaxis de la lengua literaria durante el primer español moderno (ca. 1675-1825). Martha Guzmán Riverón, Daniel M. Sáez Rivera (eds.). *Márgenes y centros en el español del siglo XVIII*. Tirant Humanidades, 203-299.
- Octavio de Toledo y Huerta 2018 = Álvaro S. Octavio de Toledo y Huerta 2018. ¿Tradiciones discursivas o tradicionalidad? ¿Gramaticalización o sintactización? Difusión y declive de las construcciones modales con infinitivo antepuesto. José L. Girón Alconchel, Francisco J. Herrero Ruiz de Loizaga, Daniel M. Sáez Rivera (eds.). *Procesos de textualización y gramaticalización en la historia del español*. Iberoamericana, Vervuert, 79-134.
- Oxygen = Syncro Soft 2002-2024. *Oxygen XML Editor*. <https://www.oxygenxml.com>.
- Ramírez Luengo 2013 = José L. Ramírez Luengo 2013. *Una descripción del español de mediados del siglo XVIII. Edición y estudio de las cartas de M. Martierena del Barranco (1757-63)*. Axac.
- Sáez Rivera 2003 = Daniel M. Sáez Rivera 2003. La duplicación de clíticos en la obra de Francisco Sobrino. *Res diachronicae* 2, 327-337. <https://resdi.net/volumen-ii>.

- Sáez Rivera 2004 = Daniel M. Sáez Rivera 2004. La lengua de Francisco Sobrino: aspectos morfo-sintácticos más relevantes. *Res diachronicae* 3, 11-38. <https://resdi.net/volumen-iii>.
- Sáez Rivera 2008 = Daniel M. Sáez Rivera 2008. Leísmo, laísmo, loísmo en el siglo XVIII en España: gramáticas y norma. Concepción Company Company, José G. Moreno de Alba (eds.). *Actas del VII Congreso internacional de historia de la lengua española. Mérida (Yucatán), 4-8 de septiembre de 2006*. Vol. 1. Arcos Libros, 1087-1104.
- Sáez Rivera & Octavio de Toledo y Huerta 2020 = Daniel M. Sáez Rivera, Álvaro S. Octavio de Toledo y Huerta (eds.) 2020. *Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva. Vientos de arrastre y de cambio en la historia del español*. Síntesis.
- Seoane & Saiz 2007 = María C. Seoane & María D. Saiz 2007. *Cuatro siglos de periodismo en España. De los avisos a los periódicos digitales*. Alianza.
- TEI = Text encoding initiative. <http://www.tei-c.org>.
- TEITOK = Maarten Janssen 2014. *TEITOK - a tokenized TEI environment*. <http://www.teitok.org>.
- TF = Texto fuente.
- TM = Texto meta.
- Transkribus* = READ-COOP 2023. *Transkribus*. <https://www.transkribus.org/>.
- Weinreich 1953 = Uriel Weinreich 1953. *Languages in contact. Findings and problems*. Mouton.
- Winter-Froemel et al. 2015 = Esme Winter-Froemel, Araceli López Serena, Álvaro S. Octavio de Toledo y Huerta, Barbara Frank-Job 2015. Diskurstraditionen, Diskurstraditionelles und Einzelsprachliches im Sprachwandel: Zur Einleitung/Tradiciones discursivas, tradicionalidad discursiva e idiomática en los procesos de cambio lingüístico. Introducción. Esme Winter-Froemel, Araceli López Serena, Álvaro S. Octavio de Toledo y Huerta, Barbara Frank-Job (eds.). *Diskurstraditionelles und Einzelsprachliches im Sprachwandel/Tradicionalidad discursiva e idiomática en los procesos de cambio lingüístico*. Narr, 1-27.