

EL CORPUS DE DOCUMENTOS ESPAÑOLES ANTERIORES A 1700 (CODEA)

Grupo de Investigación de Textos para la Historia del Español (GITHE)

Pedro Sánchez-Prieto Borja

Florentino Paredes García

Rocío Martínez Sánchez

Irene Vicente Miguel

1. Líneas de investigación y principales logros del GITHE

Las principales líneas de investigación del grupo son: 1) la edición de documentos españoles antiguos, asunto del que nos ocuparemos aquí; 2) la edición de textos literarios medievales, especialmente de textos historiográficos medievales y romanceamientos bíblicos; 3) el estudio de la historia de la lengua española, desde la historia de la escritura al léxico; 4) los trabajos de Dialectología histórica, que toman como punto de partida las fuentes documentales; 5) la metodología de la crítica textual, como desarrollo teórico de las tareas editoriales y 6) la onomástica, a partir del examen de los documentos.

Entre los principales resultados de nuestro grupo cabe destacar, en primer lugar, la colección *Textos para la historia del español*, de fuentes archivísticas, con 4 volúmenes publicados hasta hoy. Por otra parte, se ha venido desarrollando el *Corpus de Documentos Españoles Anteriores a 1700*, principal objeto de este trabajo; además, nuestro grupo ha tenido un destacado papel en la edición íntegra de la *General estoria* (10 volúmenes), ya en prensa en la Biblioteca Castro. Por último, el GITHE ha contribuido al establecimiento de un estándar en la edición de textos españoles antiguos, tanto de textos literarios como de fuentes documentales.

2. Concepción empírica y aplicaciones docentes

En el GITHE partimos de una concepción empírica de la investigación Historia de la lengua española, la cual se traslada a la docencia, puesto que tratamos de reproducir con los estudiantes, a escala reducida, el proceso e la investigación. Por ello, los alumnos de la Universidad de Alcalá tienen la posibilidad de acudir a los archivos del entorno, actividad que se lleva a cabo desde hace años con buenos resultados.

3. Textos para la historia del español

La colección *Textos para la historia del español*, de fuentes archivísticas, cuenta con 4 volúmenes publicados hasta hoy. El primer volumen apareció en el 1991, y recoge trabajos de estudiantes de la Licenciatura sobre documentos de diversos archivos (se ofrece, para cada documento, un triple acceso en forma de facsímil, transcripción paleográfica y

presentación crítica, y un estudio integral del mismo). Los siguientes volúmenes son de carácter monográfico, pues están centrados en la documentación de archivos ya concretos. De este modo, *Textos para la Historia del español II* (1994) está dedicado a la documentación medieval y de los siglos XVI y XVII del Archivo Municipal de Guadalajara. En el año 2005, Florentino Paredes publicó el tercer volumen, que contiene documentos del Archivo Municipal de Alcalá de Henares. El cuarto volumen y último hasta ahora, también de 2005, obra de Pedro Sánchez-Prieto y Ana Flores, está dedicado a los fondos de la Archivo Regional de la Comunidad de Madrid, y en particular a los de la Antigua Inclusa.

4. El *Corpus de Documentos Españoles anteriores a 1700* (CODEA)

El Corpus está formado por un total de 1500 documentos, fechados entre los siglos XII al XVII. Se ha de tener en cuenta que el CODEA es un corpus no cerrado, y aunque puedan en el futuro próximo publicarse en papel selecciones proporcionales de parte de los documentos, su versión más amplia, que será accesible electrónicamente, se irá ampliando progresivamente, por lo que en diferentes fases, según la naturaleza de los fondos que se vayan incorporando, podrán variar las proporciones entre los tipos documentales que acoge, y que, de manera operativa, dividimos en cancillerescos, municipales, eclesiásticos y privados.

En cuanto a la cronología, la mayor parte de los documentos del CODEA está comprendida entre los siglos XII y XVII. Hay que precisar que no todos estos siglos están representados de igual manera en las diferentes provincias.

Atendiendo a los lugares de emisión, contamos con documentos de todo el espacio peninsular ocupado actualmente por el español, con excepción de los territorios bilingües (Galicia, Cataluña, Valencia, Baleares), que hemos dejado fuera por plantear otros problemas específicos. Sí hemos incluido documentos del País Vasco y Navarra porque para la época elegida no contamos allí con documentación en euskera. De este modo, nuestro corpus acoge fondos de las tres provincias leonesas (León, Zamora y Salamanca), de Asturias, Cantabria, el País Vasco, Navarra, de las tres provincias aragonesas (Zaragoza, Huesca, Teruel), de La Rioja, de lo que antes se llamaba Castilla La Vieja (Palencia, Valladolid, Burgos, Soria, Segovia, Ávila), de Madrid, de la Castilla más tardíamente reconquistada (Guadalajara, Toledo, Cuenca y Ciudad Real), de Extremadura (Cáceres y Badajoz), de Murcia y de las provincias andaluzas (Jaén, Córdoba, Sevilla, Huelva, Cádiz, Málaga, Granada, Almería).

La autenticación de fecha y lugar de emisión presenta en ciertos documentos algunas dificultades particulares. Tal es el caso de las cartas nacidas en un monasterio, donde muchas veces falta el lugar de emisión; la adscripción formal y material al sello de ese centro suele resultar evidente, aparte de la presencia de determinados personajes, en particular el abad, como emisores del documento mismo. Muy raramente falta la fecha, aunque con más frecuencia ésta no es completa (puede faltar el día del mes).

En cuanto a la lengua de los documentos, se ha pretendido reflejar la diversidad dialectal del español a lo largo del tiempo, por lo que, además del castellano, están bien representados en nuestro *Corpus* los dialectos históricos, entiéndase que para la Edad Media. Y aun dentro de estos dialectos se han seleccionado los documentos de modo que puedan rastrearse diferencias internas de esas variedades. También están representadas otras variedades regionales como el castellano del País Vasco, el navarro, el extremeño, murciano y andaluz (sin entrar aquí a discutir si ha de otorgárseles el rango de dialecto).

En la documentación más antigua, la del siglo XII, y aun en la primera mitad del siglo XIII, resulta difícil deslindar latín y romance (estos documentos han planteado especiales problemas de edición). Se han incluido también documentos que contienen secciones en latín, y no sólo para la época medieval, sino también para los siglos XVI y XVII.

En cuanto a la tipología de los documentos, hemos pretendido acoger toda la variedad posible, convencidos como estamos de que sólo de este modo podía darse entrada a fenómenos que no se manifiestan en todos los registros de la lengua. Hemos incluido el llamado “documento público”, que tiene validez en todo el reino, desde los tipos más solemnes, como el privilegio rodado, a otros que, como el mandato, circulaban internamente entre los “oficiales” (funcionarios, diríamos hoy) del reino. Pero la mayoría de los documentos son los considerados “privados” por la diplomática, es decir, las cartas de compraventa y donación, o los testamentos. Este tipo evidencia mejor que el documento público la variación dialectal. Y tampoco hemos rechazado el documento no notarial, sobre todo las cartas particulares, pues éstas pueden reflejar abundantes aspectos de la oralidad.

En la elaboración de nuestro *Corpus* se han incluido sobre todo originales de los documentos, aunque no se han descartado las copias. En ello no hay inconveniente alguno, siempre que se indique con claridad (hemos señalado en el regesto la modalidad textual: copia simple, traslado, confirmación, etc.). El requisito indispensable es distinguir entre fecha del documento que se copia y fecha de la copia. Sin embargo, en algunos casos resulta difícil decidir si estamos ante el documento original o una copia, y para ello habrá que considerar aspectos diplomáticos, paleográficos, lingüísticos e históricos. De esta cuestión se dará cuenta en el campo “Observaciones” de cada documento.

Tampoco hemos rehusado incluir documentos ya publicados puesto que nuestros criterios de edición son netamente distintos de los habituales.

5. Objetivos y posibilidades didácticas del CODEA

5.1. Metodología de edición que posibilita un estudio integral del documento

En cuanto al modo de editar los fondos archivísticos, proponemos un triple acceso: en primer lugar, ha de ofrecerse el facsímil, pues sólo la consulta de éste permite comprobar lecturas dudosas y corregir otras erróneas; el siguiente paso es la transcripción paleográfica, pensada para llevar a cabo estudios gráfico-fonéticos y sobre historia de la escritura; y por último, damos la presentación crítica para hacer más legible el texto y favorecer estudios sobre las sintaxis y el léxico, así como sobre dialectología histórica, además de estudios históricos y de las mentalidades y la vida privada. El triple sistema de edición queda configurado así:

Facsímil: reproducción fotográfica en color de los documentos.

Transcripción paleográfica:

1. El desarrollo de las abreviaturas se marca entre < > (*v<ezino>*); en la presentación electrónica final esta marca puede sustituirse por cursiva (“*vezino*”).
2. Se reflejan las grafías del documento (*hauer, auer, haver; dezir, decir, dezir*).
3. Se reflejan mayúsculas y minúsculas según el uso del documento (*Rio, dios, Juan lopez*).

4. No se introduce acentuación.
5. Se refleja la puntuación del documento.

Presentación crítica:

1. Se desarrollan las abreviaturas sin dejar constancia (*vezino*).
2. Se regularizan las grafías sin trascendencia fonética (*vua* > *uva*, *ssaber* > *saber*).
3. Se regula el uso de mayúsculas y minúsculas (para marcar la sintaxis y para distinguir el nombre propio del común: *el concejo*, *don Fernando*).
4. Se introduce la tilde según las reglas académicas para marcar la prosodia antigua (med. *reína*, *vío*).
5. Mediante la puntuación se refleja la sintaxis antigua.

El juego entre facsímil y transcripción paleográfica permitirá un mejor conocimiento de la historia de la escritura en la Península Ibérica. Es el caso de un fenómeno poco conocido como la escritura de *-rr* en posición final de palabra en documentos navarros y vascos de los siglos XIII y XIV, y que podría considerarse indicio fonético (Sánchez-Prieto Borja y Torrens Álvarez, en prensa).

La historia de la lengua, por razones obvias, es seguramente el objetivo central. El estudio de la sintaxis puede verse facilitado por la inclusión de presentaciones críticas con una puntuación cuidada. A título de ejemplo, de la documentación del siglo XIII del Monasterio de Oña señalaremos el empleo de *do* (nº 0163, AHN, Clero, Burgos, carpeta 285, nº 6, Monasterio de San Salvador de Oña, del año 1237) donde encontramos: 4 *el qui ouiere un bue do bestia*, 8 *si fueren los bues do las bestias*, 8 *el qui ouiere un bue do bestia*). *Los buyes do las bestias* con el valor ‘los bueyes o las bestias’), posiblemente del vasco *edo*, conjunción disyuntiva ‘o’ (Sánchez-Prieto Borja y Torrens Álvarez, en prensa), o el empleo de *a* para el valor ‘en donde’:

Esto es el heredamiento partido: **al** molino de Mogada, una tierra que coge IX tabladas sembradura; otra tierra **a** Quissiegas, que cabe III tabladas e media; **a** Fuent Manadera, una tierra de VI tabladas; en somo de los huertos, una tierra de VI tabladas; **a** la tierra del Val, otra que coge VII tabladas; **a** Sant Yago, una tierra de VI tabladas; carrera de Muviella, una tierra de V tabladas; **al** Cascajar pora VI tabladas; **a** Ferrerueta pora V tabladas; Yaer de las Pielles, II tierras de III tabladas (nº 0156, AHN, Clero, Burgos, carpeta 286, nº 4, Monasterio de San Salvador de Oña, de 1247).

No menos interés tiene el *Corpus* para el estudio del léxico, objeto de atención de algunos trabajos de miembros del grupo. La diversidad geográfica de los documentos permite estudios de distribución areal. Como muestra, véase la de los términos para el concepto ‘parcela de terreno agrícola’:

Una tierra: uso general (León, Oviedo, Santander, Palencia, Burgos, Logroño, Valladolid, Ávila, Cáceres).

Pedaço (de tierra): Zaragoza (1519 y Terrer 1381), Salamanca (1323, 1367, 1393), pero en Salamanca, *pedaço de viña*. Ávila (1301), Aguilar (1344) y Trujillo (1335).

Pieça (de tierra): Rubielos de Mora (Teruel, 1423); Nájera (Logroño) 1406, Oña (Burgos, 1254).

Troz (de tierra, de viña, de huerto): Rubielos 1384, etc.

Quiñón: Palencia (1228), Belmonte (Oviedo, 1256), Santander (1384).

Suerte: Guadalupe, Trujillo (Cáceres 1357, 1363), Logroño 1287, 1347, *Suerte de majuelo*, Muruzábal, (Navarra 1266).

Faça: Santander, Liébana, 1384, 1466... Valladolid (1296).

Por otro lado, la variedad geográfica de los documentos revelará aspectos interesantes para la dialectología histórica. A este respecto, se ofrecen en las diapositivas dos ejemplos.

Y, por último, no es poca la información de interés que para la historia en general encierran las piezas editadas. Para el historiador el valor testimonial del documento debe superponerse a otras consideraciones. En concreto, cabe destacar el interés para la historia de las mentalidades y para el mejor conocimiento de la vida cotidiana en otras épocas. Como ejemplos, ofrecemos un testimonio de una nota de abandono de una niña en la Inclusa de Madrid o una carta privada de la condesa de Abrantes.

5.2. Explotación informática

En cuanto a la anotación mediante marcas TEI u otro sistema similar, es sabido que este proceso es inseparable de la edición filológica. Un sistema amplio de etiquetado de los documentos tiene evidentes ventajas, pues permite recuperar fácilmente toda la información anotada. Sin embargo, para no retrasar la publicación hemos creído preferible poner, por ahora, los textos sin anotar a disposición del interesado. De este modo, esperamos que, de manera inmediata, el CODEA sea accesible en la página web del GITHE www.textohispanicos.es.

Por el mismo motivo, no se incluyen ahora buscadores específicamente desarrollados para el corpus. Se trabaja ya, sin embargo, en la elaboración de herramientas de recuperación de la información que permitirán establecer diversos filtros de búsqueda a partir de los campos de la cabecera. La explotación del CODEA se beneficiará pronto de la lematización completa del texto, gracias al programa electrónico *Bconcord* desarrollado por Bautista Horcajada Diezma.

Un desarrollo ulterior de CODEA consistirá en el etiquetado del texto. En resumen, prevemos tres fases en la difusión electrónica. La primera fase proporcionaría los textos en formato plano y con un buscador general; en la segunda se ofrecería un buscador elaborado expresamente para el CODEA y en la tercera el *Corpus* aparecería completamente marcado e integrado en CHARTA.

6. Estándar para la edición de fuentes documentales hispánicas

Por otra parte, el CODEA se integrará en la red CHARTA. Esta red se ha constituido recientemente, y en ella participan grupos de investigación sobre el documento antiguo de diferentes universidades e instituciones españolas y extranjeras (véase Carrera de la Red y Gutiérrez Maté en este volumen).

Nuestros objetivos principales son dos. En primer lugar, establecer un estándar para la edición de fuentes documentales en español. Estos trabajos están concluidos, y próximamente se publicarán los *Criterios de edición de documentos hispánicos (Orígenes-Siglo XIX)*. Los criterios de la “presentación crítica” de fuentes documentales han servido de propuesta básica en la ponencia debatida en el Cilengua de la Fundación San Millán de

La Cogolla, en junio de 2007, con vistas a establecer unos criterios válidos para la edición de textos literarios medievales y clásicos. Tras la discusión, y con las modificaciones pertinentes, se aprobó un estándar de presentación gráfica para textos medievales y clásicos, que se publicará próximamente. Dicho estándar ha sido aceptado por la Real Academia Española para la elaboración del corpus del *Nuevo diccionario histórico del español*.

En segundo lugar, pretendemos elaborar un corpus integrado por distintas colecciones de fuentes documentales de España y América. La idea es que cada grupo pueda trabajar independientemente, pero compartiendo criterios de edición, objetivos y herramientas de análisis.

7. Conclusiones

Las ediciones rigurosas de los textos, hechas con una metodología adecuada, nos permiten disponer de los materiales idóneos para acceder al conocimiento de la historia de la lengua desde una perspectiva más amplia e integradora. En efecto, la triple presentación de los textos posibilita estudios de carácter interdisciplinar.

Por último, la explotación informática que integra los requisitos filológicos más exigentes con las nuevas posibilidades de acceso y gestión de los documentos, constituye un valioso recurso para el investigador, además de ser aprovechable para su uso en el aula y la transmisión de los conocimientos y del método de trabajo a los estudiantes.