

EL CORPUS DE DOCUMENTOS ESPAÑOLES ANTERIORES A 1700 (CODEA)

PEDRO SÁNCHEZ-PRieto BORJA/FLORENTINO PAREDES GARCÍA/
ROCÍO MARTÍNEZ SÁNCHEZ/RUTH MIGUEL FRANCO/
MARÍA SIMÓN PARRA/IRENE VICENTE MIGUEL
Universidad de Alcalá - GITHE

1. Introducción¹

El objetivo fundamental de los corpus consiste en la creación de un marco empírico para el estudio de una lengua representado a partir del material recogido. Los corpus lingüísticos vienen a configurarse como un método eficaz para la revisión de principios teóricos, por lo que su vinculación con la lingüística es constante. Nuestra propia experiencia nos demuestra que el trabajo con corpus da origen a nuevos modelos teóricos y metodologías, por lo que la relación teoría-práctica es bidireccional.

Es en este sentido donde se integra el *Corpus de Documentos Españoles Anteriores a 1700*, que aquí se presenta.

2. El Grupo de Investigación de Textos para la Historia del Español

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,

¹ Este trabajo se ha llevado a cabo con financiación del Ministerio de Educación y Ciencia: Proyecto "Edición y estudio lingüístico de los documentos medievales de la Catedral de Toledo" (HUM2006-04767/FILO)".

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, cuestión ésta a la que nos referiremos más adelante².

3. El *Corpus de Documentos Españoles Anteriores a 1700*

El proyecto de elaboración del *Corpus de Documentos Españoles Anteriores a 1700* (CODEA) comenzó en el año 1996 gracias a un contrato inicial de investigación entre la Universidad de Alcalá y la Real Academia Española. Actualmente, el *Corpus* cuenta con más de 1500 documentos clasificados y editados. De cada uno de ellos se ha elaborado una ficha que incluye la información más relevante como pieza archivística, sobre su elaboración (lugar, fecha y escribano) y un regesto o resumen del contenido. En cuanto a la edición propiamente dicha, se ofrece facsímil, transcripción paleográfica y presentación crítica. Las transcripciones han sido revisadas varias veces con los facsímiles (e incluso en los archivos), por lo que podemos decir que estos documentos están listos para su publicación.

3.1. CARACTERIZACIÓN

Uno de los principales problemas que se le plantean al investigador que pretende elaborar un corpus textual es el de la selección de los textos que han de integrarlo. El provecho que pueda extraerse de éste dependerá, en gran medida, de cómo se haya llevado esta selección. Ello es especialmente notorio en el caso de un corpus de documentos archivísticos, entre otras razones porque el número de piezas que han de incorporarse es notablemente mayor, dada la corta extensión de los textos. Lógicamente, la naturaleza de las piezas seleccionadas ha de estar condicionada por la finalidad del corpus. Así, un corpus que pretenda ofrecer piezas representativas de la sintaxis de la lengua de uso deberá buscar aquellos docu-

² La labor cumplida a lo largo de los últimos años se concreta en numerosas publicaciones y en la elaboración de varias tesis. Entre otros, Díaz Moreno (2000); Paredes García (2003); Sánchez-Prieto Borja (1991; 1998; 2007); Torrens Álvarez (2002; 2005; en prensa); Simón Parra y Vicente Miguel (2008); Vicente Miguel (2007) y Martínez Sánchez (en prensa).

mentos que, de un modo u otro, presenten muestras de lo que se ha llamado “oralidad elaborada”, a falta de transcripciones directas de la lengua hablada (Koch y Oesterreicher 1990, Koch 1993). La situación del CODEA es más compleja en este punto, pues el objetivo es el estudio integral del documento: no hay restricciones previas, sino que se pretende favorecer una consideración del documento archivístico no reduccionista, frente a lo que ha sido habitual. La idea básica es que el documento sólo puede ser valorado, leído, estudiado, editado y aprovechado en investigaciones ulteriores, desde una perspectiva integradora de los diferentes puntos de vista (disciplinas) que confluyen en su estudio, especialmente archivística, diplomática, paleografía, historia de la lengua e historia general. Pretendemos, pues, fomentar estudios diversos y no exclusivamente lingüísticos y filológicos, por más que la articulación del *Corpus* esté pensada para favorecer trabajos sobre la historia de la escritura, la sintaxis y la distribución del léxico.

En este sentido, se ha concedido una importancia al documento oficial, administrativo y cancilleresco, mayor de la que suele ser habitual en este tipo de corpus, donde priman los textos de redacción más suelta y alejada, relativamente, de modelos preestablecidos. Es conocida la idea de que los documentos tienen un valor escaso para la historia de la lengua por su carácter fuertemente convencional, estereotipo que se aplica sobre todo al documento público, en el sentido en el que se usa el término en diplomática. Sin embargo, el manejo de amplios fondos de esta tipología demuestra que, más allá de la parte protocolaria, estas fuentes deparan sorpresas al investigador atento en todos los campos antes señalados y en ciertos aspectos, como el de la historia de la escritura, pero también en la sintaxis, resultan imprescindibles para conocer los procesos evolutivos, al marcar muchas veces tendencias que, en mayor o menor medida, fueron seguidas en otros ámbitos de producción documental, como el de los notarios públicos, eclesiásticos o incluso, a la larga, particulares³. De este modo, el documento cancilleresco se convierte en una referencia necesaria dentro del CODEA respecto a la cual pueden medirse las adhesiones e innovaciones de otros tipos documentales. Ni siquiera hemos descartado las variaciones sobre un mismo tenor textual, a saber, el que se manifiesta en los ejemplares múltiples, como las varias copias dirigidas a diferentes consejos, con sutiles diferencias gráficas, pero también de otra naturaleza, que permiten comprender ciertos aspectos de la variación tal y como se manifestaba en el texto escrito en diferentes épocas. Tales documentos permiten estudiar el alcance de la variación en diversos planos, incluso de idiolecto de quien copia (gráfico, sintáctico, p. ej., *leísmo*, presencia y ausencia de *a* ante objeto directo, etc.).

³ Tal es el caso de la adopción de la cursividad, de determinadas características ortográficas o de la innovación léxica o sintáctica del español moderno.

Se ha de tener en cuenta que el *CODEA* es un corpus no cerrado (Caravedo 1999: 98) y, aunque en el futuro próximo puedan 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.

Los documentos proceden de diversos archivos. El grueso lo constituye la sección de Clero del Archivo Histórico Nacional en Madrid (AHN), con fondos de la práctica totalidad de las provincias españolas. El AHN es riquísimo en fondos monásticos medievales, lo que le otorga un valor impagable para la dialectología histórica. El Archivo General de Simancas cubre muy bien el espacio comprendido entre los Reyes Católicos y los últimos Austrias. A ellos se añaden los documentos redactados fuera de España, como los de la sección Galeras, en los que abundan las piezas redactadas en Génova. También cabe destacar los fondos procedentes de Toledo, de la Catedral sobre todo, que permiten una nueva valoración de la llamada “norma toledana” y del castellano alfonsí (Sánchez-Prieto Borja 2007).

Lógicamente, hemos trabajado en archivos del entorno geográfico de nuestra Universidad, empezando por el Municipal de Alcalá de Henares y siguiendo por los de Guadalajara (Archivo Municipal, Archivo Histórico Provincial) o Madrid (Archivo Regional de la Comunidad de Madrid). También hemos acudido a archivos andaluces, en concreto a los municipales de Cádiz, Córdoba, Málaga y Sevilla, pues este territorio, de reconquista relativamente reciente, está peor representado en el AHN.

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.

Precisamente, aquí puede verse el espacio geográfico de nuestros documentos. Hay que señalar, claro está, que no es lo mismo lugar de procedencia, en sentido archivístico, que lugar de emisión. El AHN acoge fondos monásticos y eclesiásticos en general, que estaban en esos centros antes de ser trasladados a Madrid gracias a la desamortización de Mendizábal. Por otra parte, no todos los documentos guardados desde antiguo en un monasterio o una catedral se habían elaborado allí, sino que podían tener diversa procedencia.

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 y 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 y Teruel), de La Rioja, de lo que antes se llamaba Castilla La Vieja (Palencia, Valladolid, Burgos, Soria, Segovia y 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 andaluzas⁴ (Jaén, Córdoba, Sevilla, Huelva, Cádiz, Málaga, Granada y 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 históricos⁵, 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).

(1) *In dei nomine. Ego garsia alvarez. & uxor mea urracha martinez. uendimus tibi martino priori de sancto* (2) *turibio. quantum habemus in sancto felices. de monte in fonte. & diuisa<m> de maria alvarez mea ger(3)mana. per uiginti morabetinis. & sumus bene pagati de precio & de roboramento. (CODEA 695. AHN, Clero, Santander, carp. 1915, nº 14 (1215), carta de venta de tierras al convento de Santo Toribio).*

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

⁴ Ciertamente que con mejor y, por fuerza, más antigua representación de las occidentales.

⁵ Así, por ejemplo, sobrepasan el centenar los documentos leoneses y aragoneses.

podía darse entrada a fenómenos que no se manifiestan en todos los registros de la lengua. Hemos incluido, de acuerdo con el planteamiento expuesto más arriba, 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 este sentido, a las divisiones “clásicas” que ha ido estableciendo la diplomática y que atienden, sobre todo, a la estructura material y formal de la pieza, puede oponerse otra seguramente más operativa, la que considera el documento según su origen. Así, distinguiríamos entre documento cancilleresco, municipal, eclesiástico (con diversos niveles) y particular.

También se incluirán documentos del registro más bajo que hasta ahora hemos encontrado y que muestra una práctica peculiar de la escritura: las notas de abandono de niños que acababan en la Inclusa de Madrid (siglos XVI y XVII; *THE IV*):

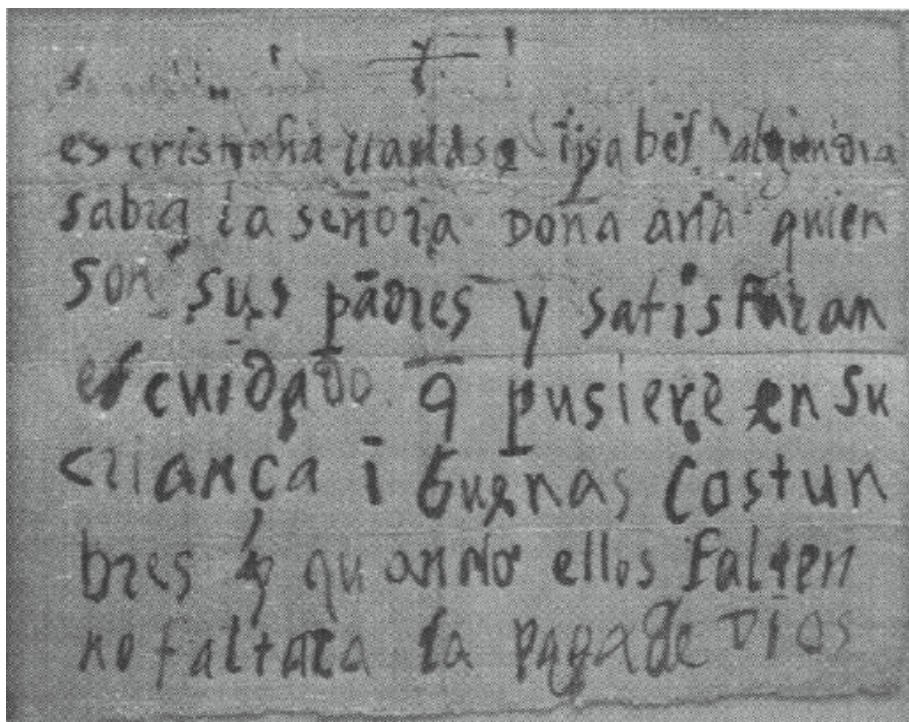
En la elaboración de nuestro *Corpus* se han incluido, sobre todo, originales de los documentos, aunque no se han descartado las copias. En ello no hay inconveniente alguno, siempre que se indique con claridad (hemos señalado en el regesto la modalidad textual: copia simple, traslado, confirmación, etc.). 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.

3.2. METODOLOGÍA: EL TRIPLE ACCESO AL 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 por último, damos la presentación crítica para hacer más legible el texto y

FIGURA 1
ARCM 8657/13, 2/36, 13 de septiembre de 1605



favorecer estudios sobre las sintaxis y el léxico (Sánchez-Prieto Borja 1998). El triple sistema de edición queda configurado así⁶:

Facsimil: reproducción fotográfica en color de los documentos.

Transcripción paleográfica:

1. El desarrollo de las abreviaturas se marca entre <> (<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.

⁶ Una formulación completa de éstos se ofrece en *Criterios de edición de documentos hispánicos (siglos XII-XVIII)*, aprobados por la red CHARTA (Corpus Hispánico y Americano en la Red: Textos Antiguos).

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.

De cada documento se ofrece una ficha catalográfica; por ahora se dará una forma provisional simplificada y en un futuro próximo se sustituirá por una versión más completa. He aquí una muestra de la versión actual⁷:

GITHE

CODEA

0981

AHN, Clero, Navarra, carpeta 1423, nº 6

1 de septiembre de 1312 (s.l.)

Carta por la que Semén Martíniz de Beortegui se compromete a pagar al Monasterio de San Pedro de Ribas dos cahíces de trigo y veinte libras de sanchetes cada año en agradecimiento por los servicios recibidos.

Pedro Julián (*escriví*)

Cristina Castillo Martínez
José Manuel Lucía Megías
Florentino Paredes García

3.3. EXPLOTACIÓN INFORMÁTICA Y DIFUSIÓN

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.textoshispanicos.es.

⁷ Contiene: 1) identificación del Grupo, Corpus y documento dentro del Corpus; 2) identificación del documento mediante siglas del archivo y signature; 3) data y localización; 4) regesto; 5) escribano y frase literal del tipo de intervención y 6) investigadores responsables de la edición.

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 Diezma⁸.

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*.

3.4. OBJETIVOS Y POSIBILIDADES DEL *CODEA*

El *Corpus* así concebido puede ofrecer a los investigadores (e incluso a los interesados por la cultura y lengua antigua, en general) muchas posibilidades de estudio. Desde luego, ha sido siempre nuestro objetivo favorecer un acercamiento integral al documento antiguo, pues la interdisciplinariedad es, en este caso, una exigencia del objeto de estudio.

Dada la variedad del *Corpus*, serán posibles estudios de tipología documental en el ámbito de la diplomática, pero también en relación con las tradiciones discursivas. El haber considerado más de una versión de un documento permite comparar el diploma genuino con su confirmación posterior, para comprobar, contra lo que suele creerse, que las modificaciones no afectan sólo a ciertos detalles gráficos o de lengua, sino a aspectos sustanciales del contenido. Es el caso, por ejemplo, del documento por el que Fernando III fija las posesiones del concejo de Alcaraz (AHN, Órdenes Militares, Uclés, caj. 365, 2), que en la versión de Alfonso X (Órdenes Militares, Uclés, caj. 213, nº 1) conoce un cambio en los lugares de referencia.

Por otra parte, el incluir facsímiles favorecerá los estudios paleográficos tanto en un plano estrictamente científico, como en el didáctico. La inclusión de diplomas cancillerescos posibilitará comprender el proceso de adopción de las innovaciones oficiales por los notarios públicos y los centros religiosos (catedrales y monasterios), proceso que no se dio al unísono en todos los lugares.

⁸ Una muestra de las aplicaciones de este programa fue presentada en el Coloquio Internacional sobre Corpus Diacrónicos en Lenguas Iberorromances celebrado en Palma de Mallorca en octubre de 2007.

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 el qui ouiere un bue do bestia* (los bues *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 tabladadas sembradura; otra tierra *a* Quissiegas, que cabe III tabladadas e media; *a* Fuent Manadera, una tierra de VI tabladadas; en somo de los huertos, una tierra de VI tabladadas; *a* la tierra del Val, otra que coge VII tabladadas; *a* Sant Yago, una tierra de VI tabladadas; carrera de Muviella, una tierra de V tabladadas; *al* Cascajar pora VI tabladadas; *a* Ferreruela pora V tabladadas; Yaer de las Pielles, II tierras de III tabladadas (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 (Vicente Miguel 2007). La diversidad geográfica de los documentos permite estudios de distribución por áreas. 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 y 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 y Oña (Burgos, 1254).

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

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

Suerte:⁹ Guadalupe, Trujillo (Cáceres 1357, 1363), Logroño 1287, 1347, *Suerte de majuelo* y Muruzábal, (Navarra 1266).

⁹ Sin duda el corpus ofrece también la posibilidad de rastrear procesos de cambio semántico, como el que, en relación a la palabra *suerte*, muestra este texto de 1266, escrito en Muru-

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

En cuanto a la onomástica, el Corpus permite conocer cómo se llevaba a cabo la denominación de las personas, así como todas aquellas fórmulas que constituían las diferentes estructuras identificativas y la variación geográfica y cronológica en el nombre de persona. Sobre esta cuestión desarrolla su tesis doctoral María Simón Parra, quien cuenta con diferentes publicaciones sobre el tema (2006).

Por otro lado, la variedad geográfica de los documentos revelará aspectos interesantes para la dialectología histórica.

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. Por ejemplo, el siglo xv se ha considerado tradicionalmente próspero para Guadalajara y su tierra (por los nobles edificios mandados construir por los Mendozas); pero en una provisión real de 1464 Enrique IV exime de pago de tributos a quienes se avecinden en Guadalajara porque “está muy despoblada” (Sánchez-Prieto Borja 1995:185-202).

En concreto, cabe destacar el interés para la historia de las mentalidades y para el mejor conocimiento de la vida cotidiana en otras épocas. Baste citar las cartas particulares del Archivo General de Simancas, donde se encontrarán detalles sobre la vida privada de la nobleza y sus intrigas políticas.

3.5. EL FUTURO DEL CODEA

A día de hoy, la principal tarea pendiente es completar algunas lagunas en sentido geográfico y cronológico que todavía presenta el *Corpus*. Creemos que es posible hacerlo con una selección de no más de 200 documentos nuevos. A esta tarea nos aprestamos en este año en curso.

Por último, en cuanto a la difusión del CODEA, nuestra intención es publicarlo en papel en su forma final. Para ello estamos ya en contacto con diversas instituciones. Pero antes queremos ofrecer ya en Internet una primera versión.

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).

zábal (Navarra): *la suert deyuso, que es de parte de buiturno, que se tiene a las viñas de la Raga, cayó por partición e por suert a doña Andrequina.*

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.

4. Conclusiones

Cabe destacar el renovado interés por las fuentes documentales para estudiar los estadios de lengua antiguos. A su vez, estos estudios cobran un alcance mayor si se integran en una perspectiva más amplia que considere el documento en todos sus aspectos, de una manera verdaderamente interdisciplinar, que abarque la diplomática, la paleografía, la historia general y, por supuesto, la historia de la lengua. Sólo de este modo es posible una comprensión global del texto, pues no es otro el objetivo último de la filología.

Por otra parte, la amplitud cronológica, geográfica y tipológica del *CODEA* favorece, unida a la masa documental, la comparación entre documentos respecto a parámetros muy diversos, que pueden combinarse entre sí. De este modo, puede estudiarse la “multivariación”, que entendemos como la interacción entre los factores que explican las características finales del texto, tales como el tiempo, el espacio, la tipología documental, el emisor, el receptor y el asunto, entre otros.

Pero para poder lograr estos objetivos (estudio integral y estudio de la multivariación) es preciso que el *Corpus* cumpla ciertos requisitos en cuanto a su elaboración. En primer lugar, que el criterio de selección de textos sea lo más amplio posible y acoja una tipología variada y, en segundo lugar, que la metodología de edición favorezca el acceso múltiple mediante una edición compleja que proporcione a) el facsímil, b) la transcripción paleográfica y c) la presentación

crítica. Y, por descontado, que los documentos sean transcritos de manera rigurosa para proporcionar textos fiables. Donde tal objetivo no sea posible, el recurso al facsímil permitirá asegurar las lecturas.

Hemos creído, pues, indispensable, cumplir los requisitos filológicos más exigentes antes de elaborar las herramientas que permitan la recuperación de la información. Éste es nuestro próximo reto, convencidos de que tales herramientas potenciarán la utilidad del CODEA. La experiencia cumplida hasta ahora nos sirve para reivindicar el papel de una filología empírica e integral como fundamento imprescindible de los nuevos desarrollos de las humanidades.

Bibliografía

- CARAVEDO, Rocío (1999): "Lingüística del Corpus. Cuestiones teórico-metodológicas aplicadas al español", en: Koch de, Josse: *Gramática española. Enseñanza e investigación. Apuntes metodológicos*. Salamanca: Universidad de Salamanca, Tomo I, vol. 6.
- DÍAZ MORENO, Rocío (2000): "Para una datación de la escritura castellana. Siglos XVI y XVII", en: *Actas I Encuentro Internacional de Filólogos Noveles, Alcalá de Henares, 15 de abril de 2000, Universidad de Alcalá-Universität Basel*. Alcalá de Henares: Centro de Estudios Cervantinos, 55-56.
- KOCH, Peter (1993): "Pour une typologie conceptionnelle et médial des plus anciens documents/monuments des langues romanes", en: Selig, Maria/Frank, Barbara/ Hartmann, Jörg (eds): *Le passage à l'écrit des langues romanes*. Tübingen: Gunter Narr, 39-82.
- KOCH, Peter/OESTERREICHER, Wulf (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Gunter Narr.
- MARTÍNEZ SÁNCHEZ, Rocío (en prensa): "Evolución discursiva del testamento: entre la tradición y la innovación", en: *VII Congreso Nacional de la AJHLE. Universidad de Alcalá de Henares, 21, 22 y 23 de marzo de 2007*.
- PAREDES GARCÍA, Florentino (2003): "Un cuadernillo del censo de vecinos de Alcalá en 1619", en: *Anales Complutenses XV*, 199-216.
- (2005): *Textos para la historia del español III. Archivo Municipal de Alcalá de Henares*. Alcalá de Henares: Universidad de Alcalá.
- SÁNCHEZ-PRieto BORJA, Pedro (coord.) (1991): *Textos para la historia del español I. Reproducción facsímil, transcripción paleográfica, presentación crítica y comentario lingüístico de documentos medievales y de los siglos XVI y XVII*. Alcalá de Henares: Universidad de Alcalá.
- (coord.) (1995): *Textos para la historia del español II. Archivo Municipal de Guadalajara*. Alcalá de Henares: Universidad de Alcalá.
- (1998): *Cómo editar los textos medievales: criterios para su presentación gráfica*. Madrid: Arco/Libros.
- (2007): "El romance en los documentos de la Catedral de Toledo (1171-1252): la escritura", en: *Revista de Filología Española*, LXXXVII, 1º enero-julio, 131-178.
- SÁNCHEZ-PRieto BORJA, Pedro/TORRENS ÁLVAREZ, Mª Jesús (en prensa): "Las tradiciones de escritura del País Vasco comparadas con las de las regiones limítrofes", en: *Oihenart*.

- SIMÓN PARRA, María (2006): *El nombre de persona en Castilla en la Edad Media*. Alcalá de Henares. Tesis de licenciatura inédita. Julio 2006.
- SIMÓN PARRA, María/VICENTE MIGUEL, Irene (2008): “Nombres de pila y apodos en la documentación toledana medieval”, en: *Nuevas perspectivas en torno a la diacronía lingüística. Actas del VI Congreso Nacional de la AJHLE (Granada, 29-31 de marzo de 2006)*. Granada: Universidad de Granada, 473-486.
- TORRENS ÁLVAREZ, M^a Jesús (2002): *Edición y estudio lingüístico del Fuero de Alcalá (Fuero Viejo)*. Alcalá de Henares: Fundación Colegio del Rey.
- (2005): “Un tipo de hipérbaton en la lengua medieval no literaria: la coordinación escindida”, en: *Actas del VI Congreso Internacional de Historia de la lengua española. Madrid, septiembre-octubre de 2003*. Madrid: Arco/Libros, II, 1169-1178.
- (en prensa): “Primeros intentos de normalización gráfica para la representación de [y] y [ž]. Usos de i, j e y”, en: *Actas del Congreso Internacional “Orígenes del español”. El Burgo de Osma, octubre de 2004*.
- VICENTE MIGUEL, Irene (2007): “El léxico de la agricultura en la documentación medieval de la catedral de Toledo (siglos XI a XIII)”, en: *Interlingüística* 17, 1058-1067.