
Un corpus documental del siglo XIX: CODOXIX

A corpus of documents of the 19th century: CODOXIX

BELÉN ALMEIDA [belen.almeida@uah.es]
Universidad de Alcalá, España

RESUMEN:

En este trabajo se presenta un corpus diacrónico en el que está trabajando un pequeño equipo de investigadoras del grupo GITHE de la Universidad de Alcalá. Se planea ilustrar el uso escrito de la lengua durante el siglo XIX por parte de personas de diversas procedencias sociales y de diferentes regiones, en principio solamente peninsulares. Los documentos se seleccionan principalmente en archivos municipales. El corpus, que estará disponible en línea, se concibe como un proyecto *in fieri* que irá creciendo anualmente hasta abarcar unos 2500 o 3000 documentos. Este corpus llevará el nombre de CODOXIX (Corpus diacrónico del siglo XIX).

PALABRAS CLAVE:

Documentos; archivos; historia de la lengua; siglo XIX; ortografía; grafía; sintaxis; léxico; nivel sociocultural; corpus

ABSTRACT:

In this article I present a new diachronic corpus on which a two-person team from the Universidad de Alcalá is currently working. The corpus is being selected in order to show the written use of Spanish in the 19th century of writers from several peninsular regions and various social and cultural backgrounds. The greatest bulk of the documents comes from town archives, as it is in them where the most usual and frequent written communications between administration and citizens are preserved. This online corpus, which will be named CODOXIX (Corpus diacrónico del siglo XIX), will lastly comprise about 2500–3000 documents.

KEY WORDS:

Documents; archives; history of the Spanish language; 19th century; orthography; graphical traits; syntax; vocabulary; sociocultural level; corpus

RECIBIDO 2014-12-24; ACEPTADO 2015-04-02



1. Presentación

En este trabajo, se presenta un nuevo corpus diacrónico concebido para ilustrar el uso escrito del español en el siglo XIX. Este es un proyecto nacido desde un primer deseo de poner a disposición de los investigadores algunos corpus pequeños que se habían ido creando para determinados trabajos (por ejemplo Almeida y Díaz Moreno 2012, Almeida 2014) y seminarios. El salto desde esta idea muy parcial hasta una concepción relativamente compleja de un corpus que abarque documentos de diversas zonas, de todo el siglo XIX y de escritores de diferentes contextos socioculturales se dio como algo natural para las dos investigadoras responsables de esta primera fase del corpus, que, como miembros del grupo GITHE (Grupo de investigación de Textos para la historia del español), tienen ya experiencia en el trabajo con el corpus diacrónico CODEA (documentos desde los inicios hasta 1700) y su ampliación CODEA+ 2015 (hasta 1800).

El siglo XIX, como se está reconociendo ampliamente desde hace ya algún tiempo y como de nuevo quedó claro con la celebración del coloquio “Una cercana diacronía opaca” en la Universidad Masaryk de Brno, resulta un momento clave en la conformación de las variedades modernas del español. Para su estudio, se dispone de una cantidad inmensa de documentación escrita, tanto impresa como manuscrita. Nuestro interés se ha dirigido a los textos manuscritos, más difíciles de investigar por falta de corpus, pero que resulta fundamental estudiar, dado que en la manuscritura se puede apreciar el uso lingüístico y gráfico de capas de población mucho más amplias que en la producción impresa, y que precisamente el siglo XIX se caracteriza por la llegada a la escritura de numerosas personas procedentes de estratos sociales que hasta entonces no habían tenido acceso a ella. La valoración de la escritura y del analfabetismo variaron enormemente a lo largo de este proceso, y aumentó bastante la necesidad de los ciudadanos de comunicarse por escrito con la administración, debido a una generalización del canal escrito con el abandono de ciertos usos orales (denuncias, quejas) que aún subsistían a principios del XIX.

El interés de los documentos manuscritos de esta centuria es por tanto muy grande para diversas disciplinas, no solo para historiadores de la lengua, sino también historiadores de la educación y de la cultura. Aunque los documentos que encontramos en los archivos son solo parte de la producción escrita, hay que decir, por una parte, que su variedad genérica y registral es mayor de lo que a veces se supone; por otra parte, que los documentos conservados en los archivos ilustran la relación entre la lengua de la administración y la lengua escrita general, la segunda notablemente influida, especialmente en su registro culto, por la primera; y por por último es necesario señalar que los archivos conservan y permiten acceder a documentos de todo tipo de escritores, mientras que el acceso a otro tipo de fondos (por ejemplo cartas, diarios u otros documentos privados) es mucho más dificultoso y, en el caso de manos menos hábiles, su misma conservación es excepcional.

El corpus que queremos realizar es un proyecto de Rocío Díaz Moreno y de la autora de estas líneas, Belén Almeida. Sin embargo, en una fase futura probablemente colaborarán en su realización otros miembros del grupo GITHE. El nuevo corpus se denomina CODOXIX, esto es, *Corpus documental del siglo XIX*. Se pretende que en unos seis o siete años llegue a incluir unos 2500 a 3000 documentos, procedentes de archivos en principio solamente peninsulares. La decisión sobre las zonas en las que se seleccionarán archivos dependerá en parte de circunstancias prácticas como posibilidades de viajar y facilidades que den los archivos en la reproducción de los fondos. Hasta ahora, hemos obtenido documentos de los archivos municipales de Alcalá de Henares, Guadalajara y Badajoz; en una primera fase se buscará seguir recopilando documentos de zonas monolingües, en archivos de Andalucía (dos), Castilla-La Mancha (otro, además de Guadalajara), Castilla-León (dos), Aragón y Murcia. Una peculiaridad es que de cada archivo se tomarán una gran cantidad de documentos, unos 250 (en archivos de ciudades muy pequeñas, probablemente menos, unos 150), por razones que explicaremos más adelante¹. Se incluirán documentos desde 1780 a 1920, para que queden enmarcados de modo más adecuado los usos que presentan los textos de las primeras y últimas décadas del XIX.

CODOXIX será un corpus online de acceso libre y gratuito. En cuanto a la presentación de los documentos, pretendemos tener una plataforma propia bastante sencilla que esperamos estará disponible en 2015, con unos 500 documentos. Podrá accederse a la transcripción completa de los documentos y, si conseguimos permisos, a la imagen. En una fase posterior del proceso, dependiendo de nuestras posibilidades técnicas y económicas y de otros factores, nos gustaría que parte del corpus pasase a formar parte de la red CHARTA.

2. La selección de documentos

Nuestro interés se dirigió desde el primer momento a ilustrar mediante el corpus el empleo de la escritura, los usos gráficos y lingüísticos, de personas de diferentes profesiones y trasfondos sociales, no solamente de profesionales de la escritura como secretarios o escribanos. Pensamos por ello en seleccionar documentos en la enorme producción de escritos creados para regular las relaciones entre la administración y los administrados. Esa es la causa de que en principio estemos buscando documentación en archivos municipales, que representan el ámbito administrativo más cercano para la mayoría de los ciudadanos, y además permiten la selección de gran cantidad de documentos debidos a escriptores de determinada zona.

En la selección de documentos hemos buscado variedad y homogeneidad: variedad para atender a la variación sociocultural, diatópica, diacrónica, y de géneros textuales, y

¹ Los documentos elegidos son generalmente breves, por tratarse de la longitud habitual de los documentos del XIX conservados en los archivos municipales, especialmente aquellos debidos a la mano de escriptores no profesionales. El corpus recogerá desde notas de 20 o 30 palabras a documentos de varias páginas.



homogeneidad para que los subcorpus correspondientes a cada archivo sean comparables entre sí.

Al acudir a cada archivo, se buscan documentos:

- idealmente, fechados y datados (aunque tenemos algunos pocos que no lo están);
- documentos con un ordenante explícito (no anónimos) y, si es posible;
- documentos en los que ordenante, firmante y escribiente sean idénticos. De ese modo, se puede relacionar la profesión, la edad y el origen geográfico de los escritores con su habilidad escritoria, sus usos gráficos, su selección léxica, etc. Esto no siempre es posible debido al amplio uso de delegación gráfica por costumbre o necesidad (v. Petrucci, 1999). Naturalmente, también la existencia de analfabetismo o de muy escasa habilidad escritoria puede ser ilustrada y estudiada mediante un corpus de estas características.

Dentro de los géneros documentales presentes en los archivos, hemos prestado especial atención a las comunicaciones de los ciudadanos con la administración, que podemos observar en peticiones de trabajo o de limosna, solicitudes de permiso de obras, ofertas de servicios, quejas, etc. También resultan de gran interés los documentos escritos por trabajadores municipales para informar de algo (denuncias, partes, informes) o solicitar algo a sus jefes (aumentos de sueldo, modificaciones de horario...).

La continuidad temática entre los diferentes archivos no es fácil, porque los archivos organizan sus fondos de acuerdo con muy diferentes criterios, y además porque los documentos archivados varían según los archivos, e incluso según el momento dentro del mismo archivo. Hay documentos que siempre se conservaron, pero otros, como notas internas o cartas de menor importancia, pueden no haberse archivado siempre, sino solo en una época y/o en un archivo; también es posible que cierto tipo de documentos resulten ser un pequeño género textual limitado a un determinado lapso de tiempo en una administración, que tuvo poco éxito y se dejó de producir pronto. Con las limitaciones impuestas por las consideraciones indicadas, se están buscando documentos en áreas temáticas como “contratación”, “solicitudes de obras”, “escuelas”, “correspondencia” (cartas recibidas), “solicitudes de trabajo”, etc., donde suelen encontrarse textos debidos a manos de diversa habilidad escritoria.

Es frecuente, y buscada, la selección en el corpus de varios documentos de un mismo tipo o de muy similares características, por ejemplo peticiones para una misma plaza de barrendero, denuncias por invadir un campo, las respuestas a una petición judicial de los alcaldes de varios pueblos, las cartas de varios presos solicitando apoyo económico... De este modo, resulta más fácil aprehender el carácter más o menos formulario de cada tipo documental, y las innovaciones personales aparecen con mayor claridad.

El alto número de documentos que se seleccionan en cada archivo es una decisión que se ha tomado teniendo en cuenta el pequeño tamaño de los documentos en general, y

también debido a que, al conformar así los documentos una suerte de “malla textual”, es posible ilustrar los usos de muchos escritores en varios documentos, a veces con años o incluso décadas de diferencia, así como, en otras ocasiones, descubrir más datos sobre escritores, ordenantes o firmantes de un documento que pueden ser mencionados también en otros.

En cuanto a la diferente habilidad de los escritores, este es un aspecto que pretendemos que el corpus ilustre de manera clara. Se están seleccionando, a este fin, documentos debidos a personas con un alto nivel socioeducacional (secretarios, jueces, arquitectos, médicos), otros que por su propio trazo muestran que sus autores tenían una escasa habilidad escritoria², y por último aquellos que, según hemos comprobado, presentan un trazo de apariencia relativamente hábil o incluso experta, pero usos gráficos propios de escritores menos avezados, como (por citar solo algunos rasgos) separación de sílabas dentro de las palabras, unión gráfica de diferentes palabras, o un uso de *h*, mayúsculas, *b/v*, etc. diferentes de los más habituales en el momento entre los escritores cultos (cf. Almeida 2014). De este tipo de escritores, con un trazo hábil pero con particularidades gráficas o de expresión, se encuentran bastantes documentos en los archivos examinados, al contrario de lo que sucede con escritores de baja habilidad (con problemas en el trazado de las letras o en su unión).

En conjunción con el dato de la edad y de la ocupación del escritor, que con frecuencia conocemos, creemos que el estudio de estos textos puede servir para conocer qué elementos eran propios de escritores cultos y cuáles fueron variando en su consideración a lo largo de la centuria. Se está intentando que en la elaboración del corpus los textos procedentes de escritores que consideramos, según las expresadas variables y otras, de nivel medio o bajo estén presentes en todas las décadas y, en lo posible, representen entre un 25 y un 50% del total de documentos. Esto, sin embargo, está resultando muy difícil para las primeras décadas del siglo XIX, hasta aproximadamente 1830 o 1835.

Otro tipo de variación que se pretende ilustrar mediante el corpus CODOXIX es la de género o sexo, aunque al realizar la selección se está comprobando que es también bastante raro encontrar escritos de mano de mujeres. Hasta ahora, hemos hallado solamente escritos autógrafos de viudas de alto nivel social, de abadesas y de maestras, y además escritos en que otras mujeres aparecen como ordenantes, en general viudas probablemente analfabetas. Ya se ha comprobado, para documentos procedentes de Alcalá de Henares, el enorme cambio sufrido en la formación de las maestras desde aproximadamente los años 30 hasta los años 80 del siglo, que se muestra en su habilidad gráfica y de expresión escrita, así como la posibilidad de que las mujeres en general presenten usos gráficos menos prestigiosos que hombres de su misma categoría social (sus maridos o hermanos) (Almeida 2014).

2 Para las características gráficas de escritores con escasa habilidad, véase Castillo (en línea), Isasi y Enríquez (2011) y Marquilhas (2000).



Aunque CODOXIX es principalmente un corpus de textos manuscritos, se ha decidido incluir también textos escritos a máquina, debido a su frecuencia en la administración desde finales del siglo XIX. Creemos que en la mayoría de los aspectos un escrito a máquina recoge el uso (orto)gráfico de modo semejante al que lo hace la manuscritura.

3. La cabecera

La cabecera que se utilizará para los documentos de CODOXIX es semejante a la que se está utilizando para el proyecto de ampliación de CODEA (CODEA+ 2015), pero se han añadido algunos campos para reforzar los aspectos que nos parecen novedosos en la selección de documentos de CODOXIX, y que pretenden sobre todo identificar y caracterizar tanto a quienes escriben los documentos como a quienes no pueden escribir y se limitan a firmarlos o a pedir que los firmen por ellos. Con estos datos, puede trazarse un perfil sociocultural de la habilidad gráfica y de la familiaridad con las normas ortográficas o con los usos gráficos o lingüísticos más prestigiosos. Prestando atención a la identidad del escritor, o, cuando no es conocida, a su letra y a sus usos gráficos, no es raro, como ya hemos dicho, que se haya comprobado que se dispone de un pequeño corpus de documentos de un mismo escritor, donde documentar la continuidad o el cambio de sus soluciones gráficas a lo largo de varios años (un ejemplo reseñado en Almeida 2014 muestra cómo desaparece en la firma de un sereno la coma entre nombre de pila y apellido en los años que transcurren entre 1851 y 1853; este uso de la coma es relativamente frecuente entre escritores menos cultos). El hecho de que un escritor sea autor de varios documentos se reflejará en la información ofrecida en la cabecera.

Una cabecera podría llegar a contener todos estos campos:

Número:

Corpus: codoxix

Archivo y signatura:

Fecha y lugar de emisión:

Escribiente 1, nombre:

Escribiente 1, ocupación:

Escribiente 1, habilidad (alta, media, baja):

Escribiente 1, edad:

(se repite tantas veces como escritores haya)

¿Hay delegación gráfica / El escribiente es distinto del firmante? (Sí/Probable/Dudoso/No)

Ordenante/s que firma/n (nombre/s, ocupación, habilidad, edad):

Ordenante/s que no firma/n (nombre/s, ocupación, edad):

Firmante diferente del ordenante y del escribiente: fórmula del firmante

Participación femenina: No / Emisora / Firmante / Escribiente

El escribiente 1 interviene también en los documentos (escribiente, firmante, ordenante):

Regesto:

Transcriptora, revisora:

Palabras claves:

Una cabecera completada tendrá este aspecto:

Número: 1

Corpus: codoxix

Archivo y signatura: AMAH, legajo 805/11

Fecha y lugar de emisión: 1851 noviembre 23 (Alcalá de Henares, Madrid, España)

Escribiente 1, nombre: Bonifacio González

Escribiente 1, ocupación: sereno

Escribiente 1, habilidad (alta, media, baja): media

Escribiente 1, edad: -

¿Hay delegación gráfica / El escribiente es distinto del firmante? No

Participación femenina: No

El escribiente 1 interviene también en los documentos (escribiente, firmante, ordenante):
codoxix 2 (escribiente), codoxix 3 (escribiente)

Regesto: Petición de adelanto de sueldo

Transcriptora; revisora: Belén Almeida; Rocío Díaz Moreno

Palabras claves: serenos, sueldo, petición, Alcalá de Henares

4. La transcripción y edición de documentos: criterios editoriales

Como resulta lógico en un corpus iniciado por dos personas que son miembros del GITHE y que han estado colaborando en CODEA, hemos partido de las normas de transcripción de CHARTA (véase la web del corpus y Sánchez-Prieto 2011).

Sin embargo, las características especiales de estos documentos y nuestros intereses nos han llevado a estudiar diferenciar en algunos puntos la presentación de los documentos de CODOXIX de la presentación propuesta por la red CHARTA para documentación del siglo XIX:

- de momento solo estamos realizando una transcripción paleográfica, no una presentación crítica, que resulta menos necesaria debido a la época del corpus;
- se reproduce la forma de las abreviaturas (letras voladas, puntos, suprrayado) sin resolverlas entre corchetes angulares, para permitir el estudio de las abreviaturas;



— también se busca reproducir mediante el tipo de letra o la organización del texto determinados elementos, como líneas libres, subrayados, tachados o tablas. En cambio no se reproducen otros rasgos como el tamaño o la inclinación de la letra, empleados también para destacar determinados elementos.

Problemas específicos de transcripción, debidos a las características de la manuscritura de este siglo, que se están presentando son, sobre todo, la dificultad para distinguir si existe coma (a veces el trazo es muy leve) o, en los escritos de escritores menos hábiles, para distinguir la coma del punto; la dificultad para distinguir el punto de la *i* del acento, así como si el acento es agudo o grave; los problemas para saber si nos hallamos ante mayúscula o minúscula (un mayor tamaño de las letras iniciales no equivale a letra mayúscula; cuando la mayúscula y la minúscula tienen una forma semejante, a veces resulta difícil apreciar la diferencia) y también para distinguir si existe o no guion a final de línea cuando se divide palabra (a veces se alarga el trazo final de la última letra, pero este uso se da en fin de línea tanto en medio de palabra como a final). Un problema de gran relevancia, dado el interés por descubrir la identidad de los escritores, es el de determinar si existe la misma o distinta letra en la firma y en el cuerpo. Un documento incluido en la página que albergará el corpus especificará de modo detallado tanto las normas seguidas en la transcripción como el modo en que se han tomado las decisiones en caso de duda.

5. Conclusión

La búsqueda, selección y transcripción de documentos del XIX presentan algunas características específicas que las diferencian del trabajo con documentos de otras épocas. Por ejemplo, la abundancia de documentación y la rapidez de lectura permiten una selección más ajustada a los criterios diacrónicos, diatópicos, y sobre todo sociales y culturales arriba apuntados. En otros aspectos, sin embargo, hemos hallado muchas concomitancias con el trabajo con documentos más antiguos: del mismo modo que con estos, se encuentran documentos muy diferentes, con grados de formulismo muy distintos, temática muy variada y que ilustran aspectos muy diversos de la actividad humana. Así se encontrarán en el corpus, entre otros, un documento en el que una maestra solicita un aumento de sueldo, una nota que recomienda a un arquitecto para una plaza, la dimisión de un inspector de matadero y mercados, una denuncia a varias personas por robar palodulce, peticiones para plazas de sereno, guarda, bombero o alguacil, y denuncias o informes sobre distintos delitos y altercados, no pocas con reproducción de discurso directo. Entre los muchos aspectos que podrían ser investigados están el avance de la grafía académica a lo largo del siglo, el uso de abreviaturas, la aparición de la oralidad en los escritos, el leísmo y laísmo, los sistemas de puntuación... Para dar algunos ejemplos concretos, una investigación llevada a cabo sobre usos gráficos de solo 200 documentos

del Archivo municipal de Alcalá de Henares (Almeida, 2014) mostró que, para el verbo *haber*, las formas con *-v-* “eran aún preponderantes en las dos primeras décadas del siglo, mientras que después se dio un periodo de convivencia en los usos cultos de formas con *-v-* y con *-b-*”, mientras que desde los años 30 del siglo “comprobamos el declive de *-v-*, con uso exclusivo de escritores medios”. En cuanto al uso, muy característico del XIX, de la *-s-* final de sílaba (ante consonante) por actual *-x-*, el uso de estos 200 documentos, seleccionados para ilustrar todas las décadas y, en la medida de lo posible, diferentes grados de habilidad escritoria y de formación, “parece sugerir que el uso de *-x-* [...] descendió durante el primer tercio del siglo, y que los escritores, tanto cultos como menos cultos, empezaron a preferir las grafías con *-s-* [...] hasta una recuperación [de *-x-*] iniciada por escritores cultos en las últimas dos décadas del siglo”.

Mediante el estudio de un corpus más amplio, como el que nos proponemos realizar y poner a disposición de los investigadores, podrán estudiarse, como ya se ha dicho, estos y otros muy diversos aspectos del uso de la escritura por los hablantes alfabetizados en el siglo XIX.

A lo largo de los próximos años confiamos en poder ir ofreciendo entregas de un corpus sencillo, pero elaborado con seriedad que esperamos que pueda contribuir a un mejor conocimiento de la lengua del XIX.

Referencias bibliográficas

- Almeida Cabrejas, B. (2014). Scriptores con bajo y medio nivel socioeducacional en documentos del siglo XIX del Archivo Municipal de Alcalá de Henares: acercamiento a sus usos gráficos. In R. Díaz Moreno, & B. Almeida (Eds.), *Estudios sobre la historia de los usos gráficos en español* (pp. 167–210). Lugo: Axac.
- Almeida, B., & Díaz Moreno, R. (2012). Estudio lingüístico de inventarios de conventos de Guadalajara en la época de la desamortización. *Revista de Historia de la Lengua Española*, 7, 39–70.
- Castillo Gómez, A. (2003). Das mãos ao arquivo. A propósito das escritas das personas comuns. *PerCursos*, 4 (1), 223–250. Disponible en: <http://www.periodicos.udesc.br/index.php/percursos/article/view/1468/1398>.
- CHARTA: Corpus Hispánico y americano en la red: textos antiguos (<http://charta.es>).
- GITHE, CODEA: Corpus de Documentos Españoles anteriores a 1700 (<http://demos.bitext.com/codea/>)
- Isasi Martínez, C., & Enríquez Fernández, J. (2011). El castellano en documentos del siglo XIX: testimonios de manos inhábiles. *Moenia*, 17, 421–434.
- Marquilhas, R. (2000). *A faculdade das letras. Leitura e escrita em Portugal no séc. XVII*. Lisboa: Imprensa Nacional-Casa da Moeda.



- Petrucci, A. (1999). Escribir para otros. In *Alfabetismo, escritura, sociedad* (pp. 105–116). Barcelona: Gedisa.
- Sánchez-Prieto, P. (2011). *La edición de textos españoles medievales y clásicos. Criterios de presentación gráfica*. San Millán de la Cogolla: Cilengua.