

Compilación del CoDiECan: Subcorpus de Viera y Clavijo

VICTORIA GALVÁN GONZÁLEZ

ELENA QUINTANA TOLEDO

Universidad de Las Palmas de Gran Canaria

Resumen

El objetivo de este trabajo es presentar el Corpus Diacrónico del Español de Canarias y, concretamente, el Subcorpus de Viera y Clavijo. Este proyecto de investigación se está desarrollando actualmente en la Universidad de Las Palmas de Gran Canaria a cargo de los miembros de la División de Tecnologías Emergentes aplicadas a la Lengua y la Literatura del Centro Tecnológico para la Innovación en Comunicaciones. Se ofrece una descripción de los textos que se están compilando en relación a los géneros que se incluyen, la localización y selección de los mismos, y las convenciones editoriales que guían la transcripción. El CoDiECan contará, además, con una herramienta informática específica que facilitará la gestión y explotación del corpus en investigación, de la que también se realizará una breve descripción.

Palabras clave: corpus diacrónico, español de Canarias, Viera y Clavijo, herramienta informática

Abstract

This paper aims at presenting the Diachronic Corpus of Canarian Spanish and, specifically, the Subcorpus of Viera y Clavijo. This project is currently underway in the University of Las Palmas de Gran Canaria in the research unit División de Tecnologías Emergentes aplicadas a la Lengua y la Literatura of the Technological Centre for Innovation in Communications. The presentation will touch upon the genres which are included, the localisation and selection of texts, and the editorial conventions used in the transcription. Researchers will be able to exploit this corpus thanks to a corpus management tool which is being implemented as well. Some notes regarding its design will be also provided.

Keywords: diachronic corpus, Canarian Spanish, Viera y Clavijo, corpus management tool

1. INTRODUCCIÓN

La creación de un corpus diacrónico de una lengua o de alguna de sus variedades es esencial para realizar análisis lingüísticos desde el punto de vista histórico. Este tipo de corpus no sólo proporciona información acerca de las peculiaridades de una lengua y su evolución en el tiempo, sino que también permite conocer aspectos contextuales necesarios para la interpretación de los datos. En el caso del español de Canarias, ya se ha hecho notar la necesidad de investigar en esta dirección. Medina López (1994-1995: 221) señala lo siguiente:

investigar los textos redactados en las Islas ayudará a situar la lengua en un contexto histórico y social determinado; aportará información sobre la fecha de algunos fenómenos como el seseo o ceceo [...] En el terreno léxico, por ejemplo, se ofrecerían datos de gran interés, sobre todo en la línea de los arcaísmos [...], al igual que la introducción de portuguesismos y occidentalismos.

El objetivo de este trabajo es mostrar la labor realizada por el momento en la compilación del *Corpus Diacrónico del Español de Canarias* (CoDiECan) y, específicamente, en el *Subcorpus de Viera y Clavijo*. Este proyecto se está desarrollando actualmente en la Universidad de Las Palmas de Gran Canaria en el seno de la División Tecnologías Emergentes aplicadas a la Lengua y la Literatura (DTELL) del Centro Tecnológico para la Innovación en Comunicaciones (CeTIC).

En general, nuestra motivación parte de la necesidad de preservar el patrimonio cultural de las Islas y nuestro propósito es crear una compilación electrónica de buena parte del material histórico que se encuentra custodiado en diferentes archivos de Canarias comenzando por la obra literaria de José de Viera y Clavijo. El CoDiECan podrá ser accesible desde un sitio web y gestionado a través de una herramienta informática específicamente diseñada para ello.

2. ANTECEDENTES DE CORPUS SIMILARES

La lingüística de corpus ha experimentado un auge considerable en las últimas décadas gracias al desarrollo de proyectos como *Survey of English Usage Corpus*, *Cobuild Corpus* o *Brown Corpus*, por mencionar algunos, favorecidos en buena medida por los avances de la tecnología computacional. En lo que a la lengua española se refiere, existen compilaciones electrónicas de gran importancia como el *Corpus Diacrónico del Español* (CORDE) (<http://corpus.rae.es/cordenet.html>). Éste es un corpus textual que contiene muestras de todas las épocas del idioma hasta el año 1975, y cuenta ya con 300 millones de palabras aproximadamente. En él se puede encontrar una amplia variedad de géneros con textos procedentes mayoritariamente de España e Hispanoamérica, organizados en tres períodos: desde los orígenes del idioma hasta 1491, desde 1492 hasta 1712, y desde 1713 hasta 1974.

En el panorama dialectal concretamente, hay que destacar el trabajo que se está llevando a cabo para la construcción de corpora electrónicos como el *Corpus Diacrónico del Español de Chile* (CorDECh) (Contreras Seitz, 2009) o el *Corpus Histórico del Español de México* (CHEM) (<http://www.iling.unam.mx/chem/>). Ambos están siendo estructurados atendiendo a cuestiones cronológicas, geográficas y genéricas, y, tal y como lo ha hecho la Real Academia Española para la computerización del CORDE, están siendo etiquetados siguiendo el estándar internacional SGML (*Standard General Markup Language*) y las recomendaciones de TEI (*Text Encoding Initiative*).

En el ámbito dialectal que pretende cubrir el CoDiECan, es preciso mencionar la labor que está desempeñando un grupo de investigadores de las Universidades de La Laguna y de Las Palmas de Gran Canaria. Dicha labor ha quedado materializada en la publicación de numerosas obras como el *Atlas Lingüístico y Etnográfico de las Islas Canarias* (ALEICan) (Alvar, 1975-1978), un punto de referencia básico que proporciona información fonética, gramatical y léxica fundamental para el desarrollo de los estudios de dialectología de esta variedad del español. En el terreno de la lexicografía, destaca el *Tesoro Lexicográfico del Español de Canarias* (Corrales Zumbado, Corbella Díaz y Álvarez Martínez, 1996), un trabajo verdaderamente novedoso en cuya edición participó la Real Academia Española y tras cuya publicación empezaron a editarse otros diccionarios como el *Diccionario de Canarismos* (Lorenzo, Morera y Ortega, 1994), el *Gran Diccionario del Habla Canaria* (O'Shanahan, 1995) y el *Diccionario Diferencial del Español de Canarias* (Corrales Zumbado, Corbella Díaz y Álvarez Martínez, 1996). Sin embargo, no existe un corpus que tenga las características que aquí se plantea, esto es, que cubra distintas etapas de la historia del español en las Islas, que esté informatizado, que sea accesible a través de Internet y que se pueda gestionar mediante software específico.

3. DESCRIPCIÓN DEL CORPUS

Un corpus que reproduzca los textos más relevantes de la historia canaria ha de incluir la obra de un clásico como José de Viera y Clavijo (1731-1813), inserto en el contexto histórico y literario de la Ilustración. Ello explica la diversidad genérico-textual de su producción y el cultivo de las formas didáctico-ensayísticas, dada la voluntad comunicativa aneja al código ilustrado. Destaca como rasgo definitorio de la escritura dieciochesca la necesidad de transferir las novedades del conocimiento, desde la ciencia a la religión, la superación de los prejuicios que lastran el avance humano, el destierro de toda suerte de supercherías y la configuración de un modelo de lector estimulado por estos intereses. De todo ello es un ejemplo señero la obra del beneditino Fray Benito Jerónimo Feijoo y Montenegro por ser el creador del ensayo moderno en España, medio eficaz en la adopción de nuevas perspectivas en la transmisión de los varios campos del saber. El espacio diacrónico al que pertenece la obra de Viera, deudor de la obra de Feijoo, ofrece amplias posibilidades para la confección de un corpus que permita investigaciones en el plano lingüístico y literario.

3.1. Géneros textuales

El subcorpus que presentamos, según la tipología al uso, tiene un carácter textual, monolingüe, general, canónico, simple y diacrónico. Se propone mostrar un volumen de textos enteros en el contexto de la obra de un autor del siglo XVIII en el ámbito canario con las consiguientes utilidades para la investigación. Los géneros que integran su obra, de acuerdo a una estructuración interna temática, pueden clasificarse del siguiente modo:

3.3.1. Prosa:

- Obra historiográfica, con la obra por la que es un clásico de las letras canarias, en la que renueva los principios de la historiografía insular, *Noticias de la historia de las Islas Canarias*, en cuatro tomos. Su edición motiva su desplazamiento a Madrid en 1770 por la imposibilidad de materializar su proyecto literario más ambicioso en las Islas. Allí publicará los cuatro tomos (1772, 1772, 1776, 1783).

- Prosa científica, que incluye diálogos (*Librito de la Doctrina Rural para que se aficionen los jóvenes al estudio de la agricultura, propio del hombre*, impreso en 1807; *Noticias de la tierra o geografía para niños*, impreso en 1807; *Tratado sobre la barrilla dispuesto en forma de diálogo*, impreso en 1810), un diccionario (*Diccionario de Historia Natural de las Islas Canarias o Índice alfabético descriptivo de sus tres reinos Animal, Vegetal y Mineral*, impreso en 1866), memorias, exámenes científicos (que redactó para la Económica de Gran Canaria, a partir de 1785, institución de la que fue su director) y una carta (*Carta filosófica sobre la Aurora Boreal observada en la ciudad de La Laguna de Tenerife en la noche del 18 de enero de 1770*).

- Literatura de viajes: *Diarios de viaje por Francia y Flandes*, que realizó entre 1777 y 1778; *Alemania e Italia*, entre 1780 y 1781; y *por el Viso de la Mancha en Ciudad Real* en 1774. En el origen están los desplazamientos del marqués de Santa Cruz, en cuya casa vivía en calidad de ayo de su hijo.

- Novela: *Vida del noticioso Jorge Sargo: novela picaresca*, escrita en sus primeros años en Tenerife, quizá hacia 1744.

- Memorias: En el marco del género autobiográfico redactó sus memorias literarias, a instancias de Juan de Sempere y Guarinos, para incluirlo en *Los mejores escritores del reinado de Carlos III*, imprescindible para el seguimiento de la producción literaria que escribió en su dilatada biografía.

- Correspondencia: Cartas familiares. En un siglo tan proclive a la comunicación abundan los epistolarios. Viera escribió numerosas cartas a distintas personalidades del

ámbito científico, literario, político, religioso o social. Relevante es la nómina de los destinatarios que incluye, todos ellos actores y testigos privilegiados de la vida cultural y literaria de la centuria.

- Periodismo manuscrito: Obras ligadas a la tertulia de Nava, de la que fue su principal artífice: *Gacetas de Daute* y *Memoriales del Síndico Personero* (compuestas durante la década de 1760).

3.3.2. Teatro

Tragedias traducidas francesas, clásicas e italianas y piezas breves. Básicamente su obra dramática está conformada por traducciones de autores de moda como *Junio Bruto*, de Voltaire; *La Mérope*, del italiano Scipion Maffei, que tradujo también Voltaire; *Los Barmecidas*, de M. de la Harpe; o *Mitridate* y *Berenice* de Jean Racine, entre otras.

3.3.3. Poesía

- Poesías originales: Cultivó varios géneros poéticos en boga durante la centuria en la línea estético-literaria del Neoclasicismo y de la Ilustración, con pervivencia del barroquismo propio de los primeros decenios del siglo XVIII en la literatura española. Así su corpus poético incluye poesía satírico-burlesca, amorosa, didascálico-científica, patriótica o épico-heroica, circunstancial y religiosa.

- Poesías traducidas: El interés de Viera por la traducción es manifiesto a tenor del corpus que recoge su obra. Se intensificó esta tarea en especial en los años finales de su existencia. Muestra una marcada preferencia por los autores franceses, pero no están ausentes los italianos, ingleses y suizos. Los asuntos elegidos están en función de las tendencias del gusto literario adscrito al código ilustrado. Destacamos sus traducciones de la *Sátira V*, sobre la nobleza, de Boileau; *Geórgicas*, a partir de la obra de Virgilio y *Los Jardines o el Arte de hermostear paisajes*, del abate Delille; *La Enriada*, de Voltaire; o *El hombre. Poema moral*, de Alexander Pope, según la traducción francesa del abate du Resnel.

- Poesías imitadas: Con este apartado nos referimos a la clasificación que lleva a cabo el autor al distinguir entre imitación y traducción, que se explica por la conocida diversidad tipológica traductora del siglo XVIII (Urzainqui, 1991). Entre otras muestras de esta actividad del autor está el poema *Los Meses. Poema en doce cantos*, escrito en 1795, a partir del poema homónimo del francés Roucher, que publicó en 1779.

3.2. Localización y selección de los textos

La producción textual de Viera se halla básicamente en las bibliotecas públicas del Archipiélago. Es cierto que también hay testimonios manuscritos en bibliotecas privadas,

pero en este apartado sólo se puede contar con los archivos particulares de los herederos de Álvarez Rixo en el Puerto de la Cruz de Tenerife o el de los herederos del marqués de Acialcázar en Las Palmas de Gran Canaria. En principio, trabajaremos en los espacios públicos por las evidentes facilidades para el proceso investigador. Hay que señalar que, aunque las fuentes que nos interesan son manuscritas y primeras ediciones, la obra del autor cuenta también con una trayectoria impresa notable, aunque todavía insuficiente. Para toda esta información es una herramienta imprescindible la obra de Agustín Millares Carlo y Manuel Hernández Suárez (1992).

Las bibliotecas insulares que albergan los textos de Viera son las siguientes:

Biblioteca Municipal de Santa Cruz de Tenerife: Cuenta con un catálogo de manuscritos que están en formato de microfilm. La obra de Viera está en entradas referidas a títulos suyos o en colecciones de fondos procedentes de bibliotecas particulares, como el de Antonio Pereira Pacheco y Ruiz o Francisco María de León. Destacamos por su valor documental los manuscritos autógrafos de las tragedias, por ejemplo, *Berenice* de Jean Racine o *La Mérope* de Maffei, entre otras; el único manuscrito de su *Diccionario de Historia Natural de las Islas Canarias*, de 1799; la *Colección de algunos opúsculos poéticos* o *Fruta del tiempo en el Parnaso*; o también, de su puño y letra, los cuatro tomos de sus *Cartas familiares escritas a varias personalidades, por sus dignidades, clases, empleos, literatura o buen carácter de amistad y virtud*.

Biblioteca de la Universidad de La Laguna (Fondo Canario): Atesora un volumen importante de manuscritos canarios y entre ellos se hallan algunos títulos de Viera. Hay un catálogo de manuscritos del fondo antiguo (Fernández Palomeque y Morales Ayala, 2002), que se puede consultar *online*. Además, durante el mes de febrero de 2010 la biblioteca ha presentado su catálogo de manuscritos digitalizados.

Biblioteca de El Museo Canario de Las Palmas de Gran Canaria: Se trata del centro de documentación bibliográfico más antiguo de las Islas. De Viera se halla aquí el mayor volumen bibliográfico. Destaca la colección de obras copiadas por Agustín Millares Torres y Juan Padilla. Entre otras “joyas” se conserva un manuscrito autógrafo de Viera de *Los Meses* con dibujos del autor.

Biblioteca de la Sociedad Económica de Amigos del País de Tenerife. Conserva todo el fondo del marqués de Nava, amigo de Viera, que incluye obras relevantes como un manuscrito del autor del poema *Las Bodas de las Plantas*.

3.3. Transcripción de los textos

La mayoría de los textos se transcribirán directamente de los archivos o, en su defecto, utilizando manuscritos microfilmados o digitalizados. La transcripción se llevará a cabo respetando en todo momento la grafía original del texto manteniendo los signos de puntuación originales y conservando las abreviaturas, la duplicación de letras, las contracciones, el uso de mayúsculas y minúsculas, y la acentuación tal y como figuran en el original.

3.4. Codificación de los textos

En cuanto a la codificación de los textos, en los archivos aparecerá un encabezado que contendrá la información bibliográfica del documento, el tamaño del archivo y el número de palabras. Las notas marginales, la numeración de las páginas, la presencia de caracteres especiales, o el uso de idiomas distintos del español como pueden ser el inglés o el portugués, vendrán indicados mediante la codificación en XML (*eXtensible Markup Language*), utilizando para ello las recomendaciones propuestas por TEI, de forma que puedan ser eventualmente volcados en la red, y que navegadores como *Internet Explorer*, *Mozilla Firefox*, *Netscape* o *Safari* puedan interpretar la codificación y mostrar los textos limpios.

4. USO DE SOFTWARE ESPECÍFICO

Actualmente, estamos trabajando en el desarrollo de una herramienta informática específica para el tratamiento y análisis de los textos, y que, como hemos apuntado anteriormente, podrá ser accesible a través de Internet. Las primeras pruebas que se han hecho con el motor de búsqueda arrojan buenos resultados, si bien aún quedan muchos aspectos por tratar, como el indexado del corpus y la velocidad de respuesta.

5. IMPLEMENTACIÓN Y EXPLOTACIÓN DEL CORPUS EN INVESTIGACIÓN

El CoDiECan resulta atractivo por múltiples razones. Obviamente, desde un punto de vista lingüístico permitiría realizar análisis diacrónicos de fenómenos propios del habla de las Islas, lo que vendría facilitado por la posibilidad de cuantificar los datos y hacerlo de forma más rápida gracias a la herramienta informática en cuestión. Permitiría, además, acceder a la información codicológica de los manuscritos y textos impresos, así como otra relativa a la encuadernación, los escribas o las afiliaciones textuales, entre otras. Dado que los textos se encontrarán transcritos en su totalidad, permitirán un estudio pormenorizado de sus

contenidos, lo que a su vez proporcionará información de tipo social, cultural e histórica, de gran importancia para los estudios pragmáticos de los documentos. Sin embargo, el uso de este corpus no quedaría restringido únicamente a lingüistas e investigadores interesados en analizar aspectos literarios, sino que otras disciplinas como la historia, la etnografía o la antropología podrían igualmente verse beneficiadas.

6. CONCLUSIÓN

El CoDiECan llena un hueco en el estudio de la variedad del español de Canarias, sobre todo, porque no existe una compilación electrónica con textos representativos de las Islas que sean susceptibles de ser manejados mediante una herramienta informática específicamente diseñada para ello. El principal objetivo de este corpus es facilitar la realización de análisis lingüísticos y literarios, aunque también permitiría el estudio de los textos desde otras disciplinas como la historia, la antropología o la etnografía. Hasta ahora se han transcrito un buen número de textos pertenecientes al Subcorpus de Viera y Clavijo, algunos de los cuales se han codificado parcialmente para realizar las primeras pruebas del motor de búsqueda. Queda ampliar la nómina de textos transcritos para este subcorpus, codificarlos en su totalidad e introducirlos en la base de datos de la herramienta informática que se está implementando.

REFERENCIAS

- Alvar, M. (1975). *Atlas lingüístico y etnográfico de las islas Canarias* (Tomo I 1975, Tomo II 1976 y Tomo III 1978). Las Palmas de Gran Canaria: Ediciones del Excelentísimo Cabildo Insular de Gran Canaria.
- Contreras Seitz, M. (2009). *Hacia la constitución de un corpus diacrónico del español de Chile*. *Revista de Lingüística Teórica y Aplicada* 47(2), 111-134.
- Corrales Zumbado, C., D. Corbella Díaz y M^a. A. Álvarez Martínez. (1992). *Tesoro lexicográfico del español de Canarias*. Madrid: Real Academia Española – Gobierno de Canarias.
- Corrales Zumbado, C., D. Corbella Díaz y M^a. A. Álvarez Martínez. (1996). *Diccionario diferencial del español de Canarias*. Arco Libros.

- Fernández Palomeque, P. y M^a. L. Morales Ayala. (2002). *Catálogo de manuscritos de la Biblioteca Universitaria de La Laguna*. La Laguna: Universidad de La Laguna.
- Lorenzo, A., M. Morera y G. Ortega. (1994). *Diccionario de canarismos*. La Laguna: Francisco Lemus (ed.).
- Medina López, J. (1994-1995). Dialectología y diacronía en el español de Canarias: Perspectivas futuras. *Revista de Filología Románica* 11-12, 217-236.
- Millares Carlo, A. y M. Hernández Suárez. (1992). *Biobibliografía de escritores canarios (Siglos XVI, XVII y XVIII)* (Tomo VI). Madrid: Ediciones del Cabildo Insular de Gran Canaria.
- O'Shanahan, A. (1995). *Gran diccionario del habla canaria. Más de 13000 voces y frases isleñas, de utilidad para propios y ajenos, recogidas de la tradición oral y escrita*. La Laguna: Centro de Cultura Popular Canaria.
- Urzainqui, I. (1991). Hacia una tipología de la traducción en el siglo XVIII: los horizontes del traductor. En M^a. L. Donaire y F. Lafarga (Eds.), *Traducción y adaptación cultural. España-Francia* (pp. 623-638). Oviedo: Universidad de Oviedo, Servicio de Publicaciones.