

“Tesoro de la lengua castellana”. *Major Authors on CD-ROM: Miguel de Cervantes*. Ed. Roberto González Echevarría. Woodbridge: Primary Source Media, 1997.

El *Tesoro de la lengua castellana* no es un producto independiente, sino que se trata de uno de los diversos textos incluidos en un CD-ROM dedicado a la obra de Miguel de Cervantes. Requisitos mínimos del sistema, un PC 486 o superior; monitor VGA o mejor; 8MB de RAM (12MB para ver las imágenes); lector CD-ROM; entre 10 y 12 MB de espacio libre en el disco duro; Windows 3.1, o una versión más alta. El programa incluye dos CD-ROM, uno con la interfaz en inglés y el otro en español, aunque la base de datos que manejan es la misma. El programa incluye también el Borland Database Engine, necesario para acceder a la información contenida en el CD-ROM.

A diferencia de otros diccionarios electrónicos (*Diccionario de la lengua española, Diccionario del uso del español, Diccionario general de la lengua española*), el *Tesoro de la lengua castellana o española* incluido en el CD-ROM *Major Authors on CD-ROM: Miguel de Cervantes* es una base de datos electrónica que permite la consulta de la totalidad del texto del Tesoro, y la de la reproducción facsimilar digitalizada del original.

El *Tesoro de la lengua castellana* puede ser consultado página a página, accediendo a la transcripción del texto y la reproducción facsimilar. Tras seleccionar en la barra de navegación el botón **Índice de materias** aparece una ventana en la que haremos clic dos veces sobre la opción **Thesaurus/Tesoro**, y desde aquí, haciendo clic sobre una de los dos volúmenes del Tesoro abriremos la ventana de consulta de la transcripción. En la parte superior de esta ventana aparece una nueva barra de navegación con tres botones que nos permitirán avanzar (**Prox**) o retroceder (**Ant**) página a página, o ir directamente a la página que deseemos (**Ir a**). El menú desplegable **Número de página relevante** sólo se activa tras efectuar una búsqueda, presentando una lista de otras páginas

en el documento con la información buscada. En la parte inferior de la ventana aparece la transcripción del texto y un enlace que abre una nueva ventana con la reproducción facsimilar de dicha página. Es también aquí donde se nos advierte del uso de [] para indicar los caracteres no reproducibles¹.

Es posible exportar a fichero la transcripción del texto y/o las imágenes. Para ello usaremos el botón **Imprimir/Guardar** de la barra de navegación y seguiremos las instrucciones ofrecidas en el correspondiente menú. El texto se guarda en formato ASCII y la imagen en formato BMP. La calidad de la imagen –en blanco y negro– es excelente, y su gran tamaño (aproximadamente 1000x1491 píxeles) permite una fácil consulta, aunque se hace preciso utilizar constantemente las barras de desplazamiento horizontal y vertical durante la lectura.

Una de las características más notables del CD-ROM es su sistema de búsquedas avanzado (Borland Database Engine) que, gracias al uso de operadores booleanos, de proximidad, comodín y de acoplamiento, permite efectuar sofisticadas consultas de la base de datos². A continuación se describen los principales operadores y sus posibilidades.

- **Operadores booleanos:** permiten búsquedas basadas en la presencia o ausencia de términos de consulta.
 - **OR:** para búsquedas que contengan una o ambas de las expresiones de consulta en cualquier lado del operador **OR**. También puede usarse con múltiples términos en paréntesis; por ejemplo, la consulta **(amor and muerte) or (amor and vida)** recogerá todos los casos que contengan ambos “amor y muerte”, todos los que contengan ambos “amor y vida” y todos los que contengan “amor”, “muerte” y “vida”.

1 En ningún lugar del CD-ROM o del manual del usuario aparece la definición de carácter no reproducible, o un listado de todos ellos. En la parte final de esta reseña se trata en detalle este problema y sus repercusiones en la calidad final de la versión electrónica del *Tesoro*.

2 Existen también operadores de fecha, de concepto y de relevancia, que no han sido incluidos en esta reseña por su poca o nula utilidad en la consulta del *Tesoro*.

- **AND:** para búsquedas que contengan ambas expresiones de consulta juntas en cualquier lado del operador **AND**. Los términos múltiples se pueden buscar con el uso de paréntesis.
- **NOT:** dependiendo del contexto dentro de la consulta, el operador **NOT** se puede usar de forma binaria o unaria. En su forma binaria, **NOT** opera en las expresiones que inmediatamente le preceden o siguen, recogiendo todos los casos que contengan la expresión que precede al operador **NOT**, pero que no contengan la expresión que sigue al operador **NOT**; la consulta **(amor and vida) not (trabajo or muerte)** recogerá todos los casos que contengan ambas “amor” y “vida” pero que no contengan ni “trabajo” ni “muerte”. En su forma unaria, **NOT** opera en las expresiones de consulta que lo siguen inmediatamente, recogiendo todas las opciones en la base de datos que no se acoplen con dicha expresión; la consulta **not (música or risa)** recogerá todos los casos que no contengan ni “música” ni “risa”.
- **Operadores de proximidad:** permiten búsquedas basadas en la localización relativa de los términos de consulta.
 - **distancia, w/n:** este operador funciona con dos expresiones de consulta, la que le precede y la que le sigue, y determina si las dos expresiones aparecen separadas por cierta distancia **n**, en el mismo orden en que aparecen en la consulta. La consulta **verdad w/3 belleza** recogerá todos los casos que contengan la palabra “verdad” a la distancia de tres palabras de “belleza”, excluyendo las palabras neutrales³.

3 En el manual del usuario que acompaña al CD-ROM existe un apéndice que lista las 268 palabras neutrales que, según se nos explica, “no pueden ser buscadas por sí mismas, pero pueden ser buscadas en la búsqueda por una frase entre comillas sencillas” (32). Estas palabras neutrales incluyen artículos definidos e indefinidos, preposiciones y conjunciones, adjetivos demostrativos, adjetivos de uso común, pronombres personales e indefinidos, adverbios de uso común y los paradigmas de los verbos *deber, decir, devenir, estar, haber, hacer, poder, ser, tener*.

- **adyacencia, adj**: este operador, equivalente a **w/1**, localiza un término o número de términos que se encuentren adyacentes y en el mismo orden con relación a otro término o términos, excluyendo las palabras neutrales.
- **cerca, near/n**: este operador determina si las dos expresiones aparecen a una distancia de **n** palabras la una de la otra, sin importar el orden y excluyendo las palabras neutrales.
- **Operadores comodín**: permiten búsquedas de variantes de palabras que aparezcan al principio, en el medio, o al final de cualquier cadena alfanumérica.
 - **carácter singular de requisito, ?**: con este operador se representa cualquier carácter singular alfanumérico. Se puede utilizar más de un operador comodín carácter singular de requisito para buscar una cadena de una extensión específica y de contenido variable. Por ejemplo, la consulta **corre?** localizará todos los casos que incluyan “correa”, “corren”, “correr”, “corres”, pero no los que incluyan “corre” o “corriendo”; la consulta **???endo** localizará todas las palabras de siete letras que terminen en “endo”.
 - **carácter singular opcional, \$**: este operador representa cero o un sólo carácter. La consulta **cocina\$** localizará todos los casos de las palabras “cocina”, “cocinar”, o “cocinas”; la consulta **co\$рто** localizará los casos que contengan tanto “corto” como “coarto”.
 - **cadena opcional de caracteres, ***: este operador localizará cualquier cadena de caracteres alfanuméricos de cero hasta 255 caracteres. La consulta **corre*** recogerá las palabras “corre”, “correré”, “correría”, “correremos”.
- **Operadores de acoplamiento**: La opción **enraizar**, activa por defecto en el programa, funciona como un lematizador, despojando a los términos de búsqueda de los sufijos de número o flexión verbal. Si la palabra consultada es **caminar**, el operador **enraizar** asegura que las variantes “camina”, “caminando”,

“camino”, etc. sean incluidas en la búsqueda. Para eliminar esta opción es necesario utilizar los siguientes dos operadores.

- **término único/acoplamiento exacto**, “ ”: al situar un término de consulta entre comillas dobles el programa busca ese término exactamente como aparece en la consulta, sin buscar derivados. La consulta “**corre**” recogerá sólo la palabra “**corre**”.
- **multi-término/acoplamiento exacto**, ‘ ’: al situar una cadena de consultas entre comillas sencillas el programa busca esa cadena exactamente como aparece en la consulta, excepto las palabras neutrales.

Como muestra de las posibilidades de consulta que ofrece el CD-ROM basten las siguientes:

- Para ver qué lemas tratan de plantas efectuamos la consulta “**planta**” (evitando con el uso del operador **término único/acoplamiento exacto** la mayoría de las formas del verbo plantar). La consulta arroja 80 casos, entre ellos *abrojo*, *adelfa*, *agárico*, *alargez*, *alcaparra*, *fumusterrae*, *girasol*, *grana*, *herbatum*, *humilde*, *jacinto*.
- La consulta “**pescado**” or “**pez**”, con los operadores **término único/acoplamiento exacto** y **or**, arroja 123 casos.
- La consulta **hacer adj amistades**, con el operador de **adyacencia** ofrece los siguientes lemas: *apaciguar* (hacer amistades entre los que están reñidos), *besar* (que los que hizieren amistades por rencillas pasadas...), *componer* (Otras vezes hazer amistades...), *confederar* (hazer amistades, y uniones)⁴.
- La consulta **capti* or cauti* or cavti***, con los operadores **or** y **cadena opcional de caracteres**, nos permite ver en qué lemas hay referencia a cautivar, cautividad, cautiverio, cautivo, consideran-

4 Esta consulta da como resultado el lema *aceite* (es enconar más las amistades), un evidente error del sistema de consulta ya que no hay forma adyacente del verbo *hacer*.

do además las diferentes variantes ortográficas. Aparecen los términos de esta consulta en 34 lemas: *águila, ahorrar, alfaqueque, apeldar, arraax, arropas, cabañuelas, cadena, calabozo, captivo, captividad, cativear, cativeo, captiverio, clavo, cocodrilo, cencho, cepo, elefante, Empurias, enechado, esclavo, esfinge, flor, galera, huirse, lengua, libertad, merced, Orán, preso, rehenes, Roma, Tarifa.*

El enorme potencial que esta versión electrónica del *Tesoro* podría tener se ve enormemente disminuida por una serie de defectos. En primer lugar, se echa en falta la posibilidad de una consulta a partir del lema. En su estado actual, la consulta por lema genera una lista de resultados con la presencia de dicha palabra en toda la base de datos, sin diferenciar lema del texto de la definición. La solución es extremadamente sencilla, pues hubiera bastado que el programa permitiera la búsqueda ateniéndose al uso de mayúsculas o minúsculas, ya que los lemas en la transcripción siempre aparecen en mayúsculas.

Este no es, sin embargo, el mayor defecto de la versión electrónica del *Tesoro*. En una obra que pretende ser “una herramienta de investigación para estudiantes e investigadores”, que permite la creación de “índices personalizados o concordancias” (menú de Ayuda), es inconcebible que la transcripción del texto esté plagada de errores. Como muestra, baste la entrada ABARCA, de la cual se ofrece la reproducción facsimilar y la transcripción que aparece en el CD-ROM⁵.

Como puede verse en la tabla 1, los problemas son de dos tipos, uno que afecta a la ortografía y otro a la división de las palabras.

Los arriba mencionados caracteres no reproducibles no son sino la ç y la ñ y las marcas de abreviación⁶. Con respecto a los dos primeros, no existe ningún problema técnico a la utilización de estos caracteres

5 La negrita y los subrayados son míos y sirven para indicar los errores.

6 Los caracteres no reproducibles incluyen también las etimologías hebreas y griegas, tan del gusto de Covarrubias. De nuevo, se trata de una dificultad inexistente, pues el sistema operativo Windows permite la utilización de diversos tipos de letra.

Tabla 1

A B A R C A, vn genero de calçado rustico de que vsan los que viuen en sierras y lugares asperos. Sō en dos maneras, vnos de palo, que por tener forma de vacas, se dixeron auarcas : y otros de cueros de vaca crudos, que con vnos cordeles se los atan a los pies, sobre vnos trapos con que huellan sin peligro la nieue. *Latinè*, pero, peronis, à **pera**, que es el zurron; porq̃ la tal auarca es como vn zurron en que se mete el pie, como **borzegui**, aliàs **bursegui**, se dixo abursa. El Rey don Sancho de Na uarra tuuo este apellido de Auarca, o por auerse criado quando niño en abito de Serrano para **estarmas** disimulado; o porque auiedo de pasar los Pirineos para ir a socorrer a Pamplona que la tenian cercada los Moros; y estando **car gados** de nieue dio orden como los **pas sassen** con estas auarcas. El linage de los Auarcas (que es verisimil decender de f. te Rey) ion muy nobles en Castilla.

ABARCA, vn genero de **calçado** rustico de que vsan los que viuen en sierras y lugares asperos. Sō en dos **ma nerar, vnos** de palo, que por tener forma de vacas, se dixeron auarcas : y otros de cueros de vaca crudos, que con vnos cordeles se los atan a los pies, sobre vnos trapos con que huellan sin peligro la nieue. *Latinè*, pero, peronis, a **pera, que** es el zurron; porq̃ la tal auarca es como vn zurron en que se mete el **pie como borzegui, aliàs bursegui**, se dixo abursa. El Rey **do** Sancho de **Na uarra** tuuo este apellido de Auarca, o por auerse criado quando **niño** en abito de Serrano para **estarmas** disimulado; o porque **auiedo** de pasar los Pirineos para ir a socorrer a Pamplona que la tenian cercada los **Moros; y** estando **car gados** de nieue dio orden como los **pas sassen** con estas **auarcas**. El linage de los **Auarcas** (que es verisimil decender deste **Rey**) ion muy nobles en Castilla.

en una base de datos electrónica, pues se hallan definidos en la tabla ASCII en las posiciones 135 y 164, respectivamente. La transcripción de las palabras *cal[]ado* y *ni[]no*, es por lo tanto incorrecta. Además, estos dos caracteres, supuestamente no reproducibles, son transcritos como tales en varias ocasiones:

PENCA, se llama el *açote* del verdugo, a forma por ser ancha, como la penca del cardo.

PE A, ... *Peñañiel* en Castilla la Vieja villa muy noble de los Duques de Osuna

Las marcas de abreviación tampoco tendrían que causar ningún problema a un transcriptor familiarizado con el español del siglo XVII, y por lo tanto, transcripciones del tipo *só/sō* por *son*, *do[]/dō* por *don*, *por[]* por *porque*, *aui[]do* por *auiendo*, *Latinè* por *Latimem*, *nōbre* por *nombre*, *Frācia* por *Francia* son totalmente inaceptables.

Por lo que respecta a la división de las palabras, si en el original no

existen guiones al final del renglón, las palabras quedan separadas: *ma neras, Na uarra, car gados, pas sassen*. Y si en el original dos palabras, o una palabra y un signo de puntuación están unidos, no se separan en la transcripción: *neras, vnos, estarmas* (estar mas), *pie, como, borzegui, aliàs, bursegui, se, Moros: y, auarcas. El, Auarcas (que, Rey)son*. De nuevo, un mediano conocimiento del español, algo dudoso en los transcriptores, habría resuelto la aparente dificultad.

Como era de esperar, estos errores de transcripción hacen que las consultas a la base de datos sean, cuando menos parciales, y no siempre acertadas. La consulta “**braço**” solamente recoge un número limitado de casos, ya que ignora las transcripciones del tipo *bra[/]o*. Por último, el mecanismo de consulta no es fiable al 100% pues, además del caso señalado de **hacer adj amistades**, con la referencia errónea a aceite, al hacer la consulta penca, que por no incluir el operador **término único/acoplamiento exacto** hace que el lematizador funcione, presenta las palabras *pena, pendencia, pensión y peña*, como resultados, erróneos todos ellos pues ninguno tiene relación morfológica/semántica con *penca*.

Para concluir, las enormes posibilidades de consulta del Borland Database Engine se ven tan seriamente limitadas por una pésima transcripción, que hacen que el uso de esta versión electrónica del *Tesoro de la lengua castellana o española* de Sebastián de Covarrubias como herramienta de investigación sea prácticamente imposible. Con este CD-ROM se ha perdido la oportunidad de hacer que una de las joyas de la lexicografía del siglo XVII sea consultada con herramientas del siglo XXI.

BIBLIOGRAFÍA

- ALVAR EZQUERRA, Manuel, ed. (1997): *Diccionario general de la lengua española VOX*. CD-ROM. Barcelona: Biblograf.
- COVARRUBIAS, Sebastián de (1943): *Tesoro de la lengua castellana o española*. Ed. Martí de Riquer. Barcelona: Horta.
- (1994): *Tesoro de la lengua castellana o española*. Eds. Felipe C. R. Maldonado y Manuel Camarero. Madrid: Castalia.

- (1998): *Tesoro de la lengua castellana o española*. Edición facsímil. Barcelona: Alta Fulla.
- MOLINER, María (1996): *Diccionario de uso del español*. CD-ROM. Madrid: Editorial Gredos.
- REAL ACADEMIA ESPAÑOLA (1995): *Diccionario de la lengua española*. CD-ROM. Madrid: Espasa Calpe.

Francisco Gago Jover (College of the Holy Cross)