

INFORMÁTICA Y LEXICOGRAFÍA: LA EXPERIENCIA DEL *DICCIONARIO GRIEGO-ESPAÑOL*

The authors expound the great advantages brought about by the introduction of data processing to the lexicographic project of the *Greek-Spanish Dictionary*.

I. INTRODUCCIÓN

Los dos últimos años han sido decisivos en la realización del *DGE*. Gracias a fondos más abundantes proporcionados por la Comisión de Investigación Científica y Técnica, ha sido posible intentar una gestión integral del diccionario con ayuda de los ordenadores. El plan era el de aplicar estos nuevos métodos a un proyecto tan vasto como el del *DGE*, que había venido realizándose desde hacía más de veinte años con métodos tradicionales.

Hace tiempo nos habíamos acercado a los ordenadores en nuestro trabajo en el diccionario al, por ejemplo, iniciar una temprana relación con el *Thesaurus Linguae Graecae (TLG)* de Irvine, California. De esta colaboración surgió, entre otras cosas, la posibilidad de disponer de concordancias de autores que carecían de ella, como Estrabón o los novelistas griegos. En segundo lugar, podemos mencionar la elaboración, gracias a la colaboración del Centro de Cálculo de la Universidad Complutense, de listados de palabras de baja frecuencia en autores tan desatendidos en los léxicos como Galeno y Pseudo Galeno, sobre la base de las cintas magnéticas del *TLG*. Esta línea de trabajo, ciertamente ardua, ha sido abandonada al poder disponer ahora del disco compacto (CDROM) del *TLG*, en el que viene recogida la parte más significativa de la literatura griega (véase más adelante). Una tercera iniciativa en este terreno fue la de conseguir de la imprenta computerizada que imprimió los dos primeros volúmenes del *DGE* listados de las citas incluidas en ellos, con el propósito de poder corregirlas fácilmente una vez

dispuestas en orden alfabético y numérico. En definitiva, digamos que esta actividad casi pionera afectó sólo a secciones periféricas de nuestra obra¹.

El propósito de este trabajo es el de explicar a los filólogos clásicos interesados en la informática y en las indudables ventajas que ésta ofrece para el trabajo en nuestras disciplinas cual es la experiencia en este terreno de un proyecto lexicográfico de gran envergadura como es el *DGE*. Tras una primera parte, de carácter más técnico, en la que viene sumariamente descrito el equipo utilizado, en sus vertientes de «hardware» y «software», pasaremos a describir el modo en que interviene la informática en las distintas tareas de que consta el diccionario, esto es, los distintos estadios en el procesamiento de la información, desde la recogida de materiales hasta la revisión final e impresión del diccionario, pasando por la informatización de las listas iniciales de obras de referencia y de los ficheros de material, la redacción de los artículos, su inclusión en una elaborada base de datos que nos permite procesar la información de varias maneras, etc.

II. PROCESO DE INFORMATIZACIÓN

A la hora de abordar por vez primera la informatización global del *DGE* se establecieron las necesidades hardware y software. Una y otra área vienen siendo desarrolladas intentando armonizar las necesidades del proyecto con sus disponibilidades.

1. *Hardware*

Dado el volumen de redactores que integran el *DGE* trabajando al mismo tiempo se optó por un sistema multipuesto. Por otro lado, dada

¹ Sobre estas iniciativas véanse los siguientes trabajos de miembros del equipo del *DGE*: J. López Facal, «Más información sobre el Diccionario Griego-Español», *EMERITA* 46, 1978, pp. 335-341; *id.*, «The use of computers in the Greek-Spanish Dictionary», en *The possibilities and limits of the computers in producing and publishing dictionaries. Proceedings of the European Science Foundation Workshop*, Pisa, 1981, *Linguistica Computazionale* 3, 1983, pp. 97-105; F. R. Adrados, «The use of computers in the *Diccionario Griego-Español*», en *Standardization in computerized Lexicography. Saarbrücken 15-17 October 1986*, Saarbrücken s.a., pp. 161-170. Sobre el cambio de rumbo del proyecto y sus primeras realizaciones tuvo ya ocasión de hablar Elvira Gangutia en una comunicación leída en el *IX Congreso de la FIEC (Pisa 24-30 VIII 1989)*, titulada «El *Diccionario Griego-Español*: gestión computerizada de un gran léxico». E. Gangutia pronunció también una conferencia sobre este tema en la Cátedra de Lenguas Extranjeras de la Academia de Ciencias de Moscú en abril de 1990.

la especificidad del proyecto, se incorporaron equipos de alta flexibilidad como son los ordenadores personales, concretamente compatibles IBM y basados en el sistema operativo DOS de Microsoft Corporation. Las características de estos equipos son las siguientes:

a) *Unidad central*: Las unidades utilizan el procesador Intel 80286, es decir, son AT compatibles. La velocidad de proceso es excelente para la ejecución del software de tratamiento de textos que supone buena parte del trabajo a desarrollar por los redactores, y se considera buena para la ejecución de software de gestión de bases de datos. En cualquier caso el equipo se completa con una unidad basada en el procesador Intel 80386, muy indicado para la realización de tareas más complejas o lentas.

b) *Controladora de video y pantalla*: Se eligió una tarjeta EGA (Enhanced Graphics Adapter) adaptada a un monitor multifrecuencia NEC modelo Multisync II. La calidad del texto en pantalla es, creemos, lo suficientemente buena como para permitir la redacción de textos sin excesiva fatiga para el redactor. Por otro lado, la característica de tarjetas de video como la EGA o superiores que almacenan los caracteres ASCII en la memoria de la misma nos ha permitido elaborar una tabla ASCII modificada con los caracteres griegos y de otro tipo necesarios para la elaboración del diccionario.

c) *Teclado*: Español tipo «QWERTY» modificado por software para permitir acceder a otro teclado de griego, con sus signos de acentuación, diseñado específicamente para nuestras necesidades.

d) *Impresoras*: Actualmente se utilizan impresoras de matriz de puntos de 24 agujas. Estas impresoras permiten la modificación por software de los caracteres a imprimir de forma que se correspondan con los visualizados en la pantalla. Asimismo, proporcionan una calidad suficiente como para imprimir claramente todos los signos de acentuación especiales del griego clásico.

e) *Red de área local*: Este dispositivo permite una gran flexibilidad en el trabajo de los redactores, al hacer posible el acceso a la información desde cualquier puesto de la red. El software de la red es el mismo suministrado con el hardware (*Net30*, Invisible software). La red permite la utilización de las unidades de disco de todos los puestos, la realización de copias de seguridad o el acceso a las bases de datos y otros programas, así como a las impresoras, desde cualquier puesto. Aunque la velocidad de transmisión de datos a través de la red se considera muy aceptable, resulta conveniente utilizar el puesto adecuado cuando se

realizan tareas que requieren un acceso elevado a disco como es el caso por ejemplo de la gestión de bases de datos.

f) *Unidad lectora de CDROM*: Esta unidad (Philips CM100) permite, con ayuda del software adecuado, la explotación del CDROM C del *TLG* y de otros discos compactos con textos antiguos, como son los CDROM PHI 1 y 2 de la Packard Humanities Foundation. El acceso a esta información puede efectuarse desde cualquier puesto de la red como si se tratase de una unidad de disco más.

2. *Software*

La mayor parte del software utilizado en el proyecto es o está basado en software comercial. Hemos desarrollado también software «a medida» para facilitar el procesamiento de la información:

a) *Proceso de textos*: Actualmente utilizamos un procesador ampliamente difundido como es *WordPerfect* (WordPerfect Corporation). La gran difusión de este software permite una gran versatilidad y facilidad en el manejo y transporte de la información. El aprendizaje por parte de los redactores ha sido gradual y, hoy por hoy, plenamente satisfactorio.

b) *Gestión de bases de datos*: De igual modo se eligió un gestor de bases de datos de tipo relacional ampliamente difundido en el mercado y que, aunque con ciertas limitaciones, permite una gran facilidad de transporte y crecimiento futuro. De hecho existe mucho software comercial disponible que aumenta las posibilidades ulteriores del software desarrollado en el proyecto. El programa creado funciona actualmente bajo la versión 2.10 de *Foxbase Plus* (Fox Software). Por otra parte, hemos empezado a utilizar también la base de datos *Data Perfect* de Word Perfect Corporation, muy adecuada para determinado tipo de información y de muy cómodo manejo por su total compatibilidad con el procesador de textos *Word Perfect*.

c) *Manejo de los CDROM*: Por lo que se refiere al manejo de los discos compactos (CDROM) con textos antiguos, utilizamos el programa *Searcher*, obra de R. M. Smith y otros profesores del departamento de Clásicas de la Universidad de California en Santa Bárbara. En la *Newsletter* 16 (diciembre de 1989) del *TLG* se informa sobre el *software* existente para acceder a los distintos CDROM, tanto para entorno Macintosh como para ordenadores compatibles. A juzgar por lo que allí se dice y por lo que sabemos de algunos de los programas allí menciona-

dos², *Searcher* es sin lugar a dudas el programa más completo que existe actualmente, en particular por su manejo del *Index* del CDROM C (véase más abajo).

d) *Software accesorio*: Hemos desarrollado diversos programas, fundamentalmente en lenguaje PASCAL, que permiten la conversión de la información entre los diferentes formatos utilizados en el *DGE* (formato BETA del *TLG*, formato WordPerfect del *DGE*, formato empleado por la imprenta que compone el diccionario). También hemos adaptado varios programas que facilitan el manejo de diferentes dispositivos, como el teclado, la pantalla o la impresora.

III. PROCESAMIENTO DE LA INFORMACIÓN

La realización de un gran diccionario de autoridades como es el *DGE* consta de una serie de distintos trabajos, en cierto modo encadenados y dependientes unos de otros. En el eslabón inferior de la cadena se sitúa el trabajo de documentación. Los libros (ediciones de autores, léxicos, nuevas colecciones de papiros e inscripciones, estudios de diverso tipo) se localizan en los repertorios, revistas especializadas, etc. Cuando se puede disponer de ellos, los libros son estudiados con vistas a su incorporación a las listas iniciales del diccionario y en todo caso a su aprovechamiento en el cuerpo del diccionario. Las ediciones seleccionadas y los estudios de interés para el diccionario son «despojados», esto es, en la jerga lexicográfica, se recogen en ellos referencias para el diccionario, convenientemente estudiadas. Esta información se almacena en ficheros y posteriormente es utilizada, junto a otro material de referencia, a la hora de redactar los artículos. Una vez redactados, los artículos son objeto de sucesivas revisiones de muy diverso tipo, tanto sobre el manuscrito original como sobre las pruebas de imprenta. En todos los eslabones de esta cadena, muy sumariamente descrita, interviene hoy por hoy la informática, de modo más o menos decisivo. Vayamos por partes.

² Miembros del equipo del *DGE* han tenido ocasión de probar los programas *Pandora* y *SNS-greek*, ambos para ordenadores Macintosh y el programa *LBase*, para ordenadores compatibles.

III.1. *Recogida de materiales*

a) *Consideraciones previas*

El aprovechamiento para el *DGE* del material contenido en el CDROM C del *TLG* plantea un problema básico, sea cual sea el objeto de la consulta. Las ediciones escogidas por el *TLG* para cada autor, siguiendo las recomendaciones de un comité de la *American Philological Association* encargado de esta tarea, difieren en un número importante de casos de las ediciones adoptadas por el *DGE*, en un porcentaje mucho más alto de lo que en principio se podría imaginar³. Los criterios para seleccionar las ediciones, aun siendo parecidos, no son los mismos. Por poner un ejemplo lo suficientemente demostrativo, para la mayor parte de las obras de Hipócrates el *TLG* ha optado por la vieja edición de Littré, atendiendo sin duda al criterio de la uniformidad, esto es, que la mayor parte de las obras estén contenidas en un mismo *corpus* de fácil acceso, en detrimento de numerosas ediciones posteriores y que ofrecen un texto mejor. Por otra parte, la manera de catalogar los autores y textos, ya se trate de la misma edición o de una distinta, también difiere en bastantes casos. Así, tal obra que para el *TLG* pertenece a Pseudo-Galeno para nosotros es de Porfirio. Las obras que el *TLG* recoge en una entrada *Anonymi Grammatici* el *DGE* las incluye en su lista de papiros como *Gramm. Pap.*, etc. De otro lado, el *TLG* incluye en su elenco de autores y obras un buen número de entradas que nosotros conscientemente hemos desechado, como es el caso no sólo de autores del s. VII y siguientes, sino el de, por ejemplo, ediciones de fragmentos que no recogen en ningún caso citas literales del autor en cuestión. Así sucede con algunas obras de Yámblico, Teofrasto y otros. También por ejemplo, en un autor tan problemático como Filodemo, se recogen en el *Canon* numerosas entradas que no son sino artículos de revista que estudian tal o cual pasaje, eventualmente mejorando alguna lectura. El *DGE* dispone de otras maneras de proceder más flexibles cuando hay que citar este tipo de textos recogidos en publicaciones aisladas, etc. Por otra parte, en bastantes casos el *Canon* rompe con un principio que el *DGE* sigue siempre a rajatabla, que es el de utilizar una sola edición para cada obra. Una breve justificación de esta manera de proceder se

³ Las ediciones vienen recogidas en el *Canon of Greek Authors and Works*, Second Edition, Oxford University Press, 1986, de Luci Berkowitz y Karl A. Squitier. Una lista de *Addenda et Corrigenda* al *Canon* está a disposición de los interesados (cf. la *Newsletter* n.º 10, julio de 1986). Por otra parte, es inminente la publicación de la tercera edición del *Canon* (cf. la *Newsletter* n.º 17, julio de 1990).

encuentra en la introducción del *Canon*, p. XXIII. Evidentemente, los propósitos de una y otra lista son distintos.

Todos estos inconvenientes que hemos ido viendo nos han convencido de la necesidad de disponer de unas concordancias entre ambas listas. El primer problema que ha habido que superar a este respecto es que el CDROM no viene acompañado de una lista exhaustiva de los autores que incluye, con todos los datos de las ediciones. Sí, incluye, en cambio, una versión informática, ligeramente actualizada, del *Canon*, así como una escueta lista de los autores (no las obras) recogidos en él. Partiendo de la versión informática del *Canon*, convertida, no sin alguna dificultad técnica, al formato de Word Perfect, hemos sacado listas con la siguiente información: 1. Autores y obras contenidos en el banco de datos del *TLG*. 2. Autores y obras cuya inclusión en el banco de datos está prevista, pero que aún no están recogidos. 3. Autores y obras contenidos en el CDROM C. 4. Autores y obras en el banco de datos que no figuran en el CDROM C (en general porque son textos que no han sido revisados todavía). 5. Lista de autores del *Canon* por orden cronológico. 6. Diversas listas menores. La mayor parte de estas listas las hemos hecho por orden alfabético y por orden de número, según la numeración convencional adoptada por el *TLG*. El propósito de las concordancias entre la lista del *TLG* y la del *DGE* (este trabajo está actualmente en curso) es fundamentalmente el de establecer dos categorías de obras dentro del material incluido en el CDROM, por un lado las obras de consulta directa (para las que no es imprescindible comprobar el pasaje en el libro⁴) y por otro las obras cuyas citas es preciso comprobar en nuestra edición. Se trata en general de saber siempre a qué atenernos con cualquier cita que obtengamos del CDROM.

b) *Búsquedas en el CDROM*

Por lo que se refiere a la búsqueda de palabras en el CDROM C, el software utilizado, junto a algunos programas complementarios desarrollados por nosotros, nos permite hacer los siguientes tipos de búsquedas: 1. Una o más palabras. 2. Dos palabras en proximidad. 3. Una palabra cuando no está en proximidad de otra palabra. 4. Una palabra cuando está en proximidad de una segunda palabra o de una tercera palabra. En todos los casos, «palabra» vale tanto como «patrón de pa-

⁴ En cualquier caso, siempre es un inconveniente el consultar un pasaje en el CDROM sin poder al mismo tiempo tener a la vista su aparato crítico. Conviene decir que el *TLG* no ha renunciado a añadir en algún momento a los textos su aparato crítico (cf. *Newsletter* n.º 10, julio de 1986).

labra». En las distintas búsquedas combinadas, la mayor o menor proximidad entre las palabras es fijada por el usuario, en número de caracteres o de palabras. En todos los casos, es posible el empleo en la secuencia de búsqueda de caracteres comodín (que valen por cualquier carácter) y de «sets de caracteres», esto es de varios caracteres especificados por el usuario como posibles alternativas en una misma posición de la secuencia a buscar. Todas las búsquedas pueden hacerse especificando que sólo se localice la secuencia cuando sea inicial de palabra o también dentro de una palabra. El programa permite asimismo buscar sólo nombres propios y también secuencias en final de palabra.

Las búsquedas pueden hacerse: 1. En un solo autor. 2. En una obra de un autor. 3. En varios autores o en varias obras de varios autores al tiempo (sucesivamente). Aprovechando esta última posibilidad, que es absolutamente flexible y admite cualquier agrupación de autores para buscar en ellos en la misma sesión, y valiéndonos del trabajo realizado sobre la versión informática del *Canon* contenida en el CDROM (v. *supra*), hemos preparado una serie de agrupaciones de autores para hacer búsquedas simultáneamente en ellos. Hemos agrupado los autores de dos modos: a) por géneros (comedia, tragedia, medicina, filosofía, etc.); b) por siglos (v a. C., II d. C., etc.), independientemente del género. Eventualmente, podrían hacerse agrupaciones por siglos y géneros (e. g. médicos del II d. C., etc.) y cualquier otra que pueda resultar de interés para documentar una palabra en tal o cual grupo de autores. Este tipo de búsquedas, así como también las búsquedas en todo el CDROM, también posibles aunque algo lentas, pierden parte de su interés frente a las búsquedas a través del *Index* del CDROM (v. *infra*).

La información sobre las sucesivas búsquedas que se van realizando (los autores, las palabras y las condiciones de búsqueda) se almacena en ficheros que se pueden recuperar en cualquier momento, para volver a ejecutar esas búsquedas por ejemplo con alguna pequeña variación, y en general para ganar tiempo si se desea trabajar repetidamente con los mismos autores o las mismas palabras.

Los resultados de las búsquedas (la palabra buscada, los pasajes en que aparece y un contexto de un tamaño a definir por el usuario), se graban en un fichero *ASCII* que puede imprimirse cómodamente. El texto griego aparece transcrito en el formato *BETA*, que es el formato convencional en el que está escrita la información del CDROM. Por nuestra parte, hemos desarrollado un programa para convertir estos ficheros del formato *BETA* al formato *WordPerfect* del *DGE*, de tal modo que la información pueda aprovecharse directamente por los re-

dactores del *DGE* al componer los artículos del diccionario en el procesador de textos, o bien pueda imprimirse, igualmente en griego.

Pero sin duda, lo más interesante que ofrece el programa para el trabajo lexicográfico es la posibilidad de hacer búsquedas a través del *Index* del CDROM. Este *Index* es un índice alfabético (sin lematizar) de todas las formas diferentes que aparecen en los textos contenidos en el CDROM, que suman algo más de cuarenta millones de palabras. Junto a la forma figura el número de veces que aparece. El programa permite editar el índice, desplazarse a tal o cual sección del mismo y, con un solo golpe de tecla, conocer los autores y obras en que aparece una forma. En este punto, se puede grabar esa información (tales formas asociadas a tales autores y obras) como condición para una búsqueda ulterior. La información se graba en un fichero con un nombre fijo que se recupera antes de entrar en el menú de buscar. La información puede referirse a tantas palabras y pasajes como se quiera. Asimismo, del fichero se pueden eliminar autores y obras que no se desea investigar y, una vez en el menú de buscar con todas las búsquedas preparadas para ser efectuadas una tras otra, se puede revisar cada condición de búsqueda y hacer los cambios que se estimen oportunos.

Aprovechando esta posibilidad, hemos planteado la recogida de materiales del siguiente modo. Primero llevamos a cabo un trabajo de documentación sobre el papel, comparando el índice impreso⁵ con diversos diccionarios (*LSJ*, *Lampe*, *Pape-Benseler*, *Stephanus*, etc.) y con nuestros ficheros y luego una persona se dedica a recoger en el CDROM información sobre las palabras que previamente hemos considerado interesantes. Esta información, como decíamos más arriba, es utilizada por el redactor en el momento de dar forma a los artículos en el procesador de textos. Conviene poner de relieve que la posibilidad de disponer con relativa facilidad de una masa de materiales tan ingente obliga a los miembros del equipo encargados de recoger y estudiar ese material a un esfuerzo suplementario de crítica para ser capaces de distinguir lo relevante de lo que carece de valor⁶.

En resumen, con ayuda de este programa y de algún otro desarrolla-

⁵ Aprovechamos la oportunidad para dar las gracias a A. Bozzi y A. Sapuppo, del *Istituto di Linguistica Computazionale* de Pisa, cuya ayuda para imprimir el índice ha sido decisiva.

⁶ A idéntica conclusión llega el papirologo L. Koenen en su comunicación al *IX Congreso de la FIEC (Pisa 24-30 VIII 1989)*, titulada «The TLG at Irvine and the Computerization of Subliterary Texts on Papyri», leída en la sección del congreso llamada *Instrumenta Studiorum*. En el libro de resúmenes de dicha sección, p. 70, Koenen escribe a propósito de esta cuestión: «More critical acumen is needed, not less».

do por nosotros, podemos hacer lo siguiente: 1. Realizar búsquedas sencillas o sofisticadas en autores o grupos de autores definidos por nosotros. 2. Editar cualquier sección del CDROM. 3. Hacer búsquedas muy rápidas en todo el CDROM a partir de su índice alfabético de formas, con la seguridad de que lo que se le pide lo va a encontrar. 4. Trabajar con los resultados de las búsquedas en griego en el procesador de textos en el que redactamos el diccionario e imprimirlos en griego en una impresora de 24 agujas.

III.2. *Material de referencia*

El material de referencia extraído del despojo de ediciones, colecciones de papiros e inscripciones, léxicos, índices, concordancias, artículos de revistas, etc., se almacena en una base de datos, que viene a sustituir a los ficheros repletos de fichas. Esta base de datos tiene una estructura menos compleja que la gran base de datos del diccionario (v. *infra*) y permite la introducción y almacenamiento de nuevo material para ser utilizado por los redactores. Éstos pueden consultarla cómodamente y también extraer el material deseado en forma de fichero de ordenador para utilizarlo desde el procesador de textos en el momento de redactar el artículo o para mandarlo directamente a la impresora. En el curso de un año y medio o dos años, esta base de datos ha registrado la inclusión de unas cuatro mil entradas y está en constante aumento.

III.3. *Listas iniciales*

Las listas iniciales de autores, papiros, inscripciones y abreviaturas, en las que vienen detalladas las ediciones utilizadas así como las abreviaturas correctas de los autores, obras y en general de la bibliografía citada en el cuerpo del diccionario, han sido copiadas en *Word Perfect* y trasvasadas a bases de datos individualizadas, creadas con el programa *Data Perfect*. Estas bases de datos no sólo sirven para consultar esa información cómodamente en el ordenador, sino que están preparadas para acoger información de uso interno referente a los libros y textos contenidos en las listas (modo de citar, estado de «despojo», datos del *TLG*, de otros diccionarios, cambios de edición, traducciones, índices, léxicos, concordancias, bibliografía, observaciones, etc.). El trabajo de documentación para incluir este tipo de información en las distintas bases de datos se encuentra actualmente en curso.

III.4. *Redacción*

La redacción de un artículo del diccionario, especialmente si es complejo, resulta mucho más sencilla sobre un procesador de textos que directamente sobre la base de datos. El redactor se sirve para componer sus artículos, además de las fichas en papel y de la consulta de libros de referencia, de la información, en forma de fichero en formato de *Word Perfect*, procedente del despojo del CDROM C así como de la que procede de la base de datos del material. Además, el redactor tiene la posibilidad de consultar el CDROM (búsquedas, visualización, recogida y comprobación de citas en el momento de redactar los artículos) y las distintas bases de datos (del diccionario, de material, de las listas de autores, etc.).

III.5. *Intercambio de Información*

Actualmente se está potenciando el acceso múltiple a la información de manera que desde el puesto de trabajo el redactor pueda acceder simultáneamente a la información procedente del *TLG*, al material nuevo recogido por los colaboradores en la base de datos creada a tal efecto, a la base de datos del diccionario, a las bases de datos documentales de las listas iniciales y al procesador de textos. La idea es la de poder acceder a toda la información disponible desde la pantalla del ordenador. La posibilidad de pasar de unos programas a otros sin abandonarlos previamente, es decir, acceder a ellos en el mismo punto donde se dejaron, y sin saturar la limitada memoria que permite el MS-DOS, se hace realidad mediante programas comerciales de bajo coste (*Switch-it*, *Better Software Technology*), con un rendimiento sorprendentemente alto si se aprovechan las posibilidades de memoria y velocidad de los equipos.

III.6. *Inclusión en base de datos del DGE. Utilidades de la misma*

La base de datos principal del diccionario, que hemos denominado *LABRIS*, es el programa más complejo y elaborado desarrollado en el seno del *DGE*. Su diseño inicial parte de la cooperación entre Elvira Gangutia, codirectora del proyecto, y Francisco Fernández Izquierdo, Colaborador Científico del Centro de Estudios Históricos del C. S. I. C. Este programa está registrado en el C. S. I. C. Su propósito principal

es el de recoger las entradas del diccionario en una estructura formalizada de una manera rigurosa y al tiempo flexible, capaz de procesar la información de diversas maneras según las necesidades del momento.

Posee una estructura similar a la que se observa en cualquier artículo del *DGE*. Es una estructura en árbol que permite cinco subniveles de los cuales pueden utilizarse todos o sólo algunos, dependiendo de la complejidad del artículo en cuestión. La interrelación adecuada de los diferentes niveles se realiza mediante códigos internos de forma que una vez reclamado un lema, éste se recompone buscando en los diferentes subniveles hasta completar el lema. De esta forma, existen nueve ficheros principales interrelacionados, ordenados mediante trece índices. Accesorariamente existen otros ficheros e índices.

El programa está dividido en cuatro módulos principales, cada uno de ellos estructurado por medio de procedimientos: Introducción de lemas, Modificación de lemas, Impresión de lemas y Utilidades. Este último módulo engloba la creación de índices, información estadística, enlace con otras bases de datos, etc.

Desde el punto de vista semántico, cuatro campos o áreas jerárquicamente organizados son contemplados en nuestro diccionario. Son llamadas *área general*, *área específica*, *área traducida* y *área matizada*, y se corresponden con las subdivisiones principales de un artículo del diccionario. Si tomamos, a modo de ejemplo, la estructura de la palabra βίος, que vendrá recogida en el cuarto volumen del *DGE*, observamos que tiene tres áreas generales, señaladas por letras mayúsculas y dando pie a una serie de lo que llamamos etiquetas: *A en relación con los medios económicos*; *B no directamente relacionado con los medios económicos*; *C identificado con sujeto y lugar*. Si tomamos uno de ellos, bajo el área general B hay varias áreas específicas, que en el diccionario son señaladas con números romanos: *I con varias calificaciones*; *II como algo que sucede a lo largo del tiempo*; *III opuesto a la muerte*. Bajo el área específica B III encontramos varias áreas traducidas o traducciones que serán señaladas con números árabes o simplemente con traducciones en itálica (en el caso de artículos sin grandes subdivisiones): *1 vida, existencia*; *2 vida orgánica*. Tras el área traducida o la traducción aparece un área menos formalizada que llamamos *área matizada*; aquí se recogen una tras otra explicaciones, acepciones, etc., que ayudan a precisar la traducción en itálica o parten de ella. Este área matizada no viene señalada con ningún signo en particular: sólo en ocasiones en palabras muy largas la señalamos con letras minúsculas.

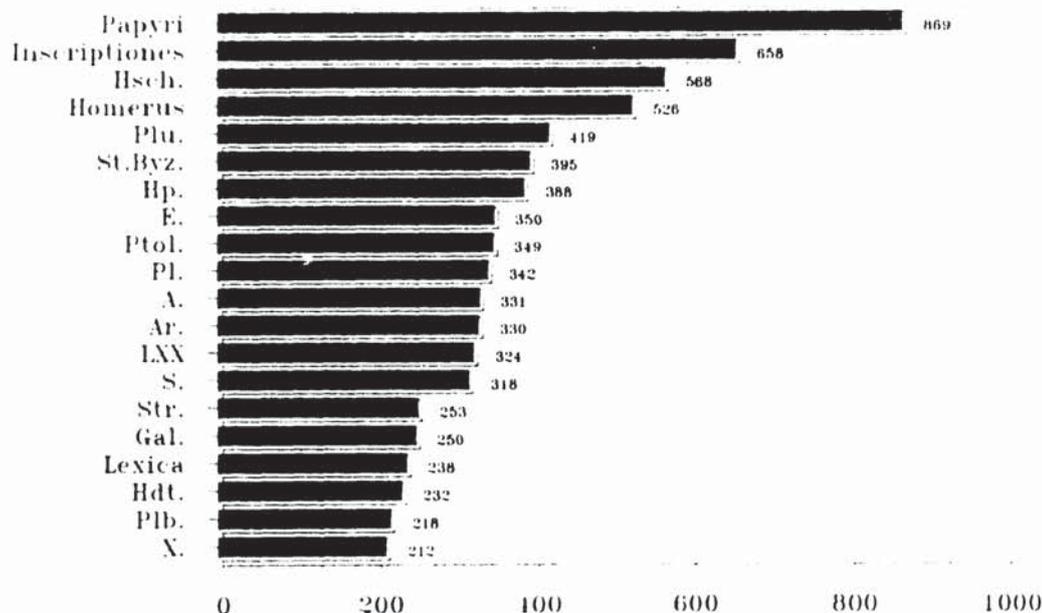
Desde un punto de vista formal el programa va registrando una tras otra todas las abreviaturas y los nombres completos de los autores anti-

guos y sus obras, de tal modo que reconoce las abreviaturas ya introducidas y detecta los errores. No hemos incluido estos datos desde el primer momento, y el inventario se ha ido haciendo progresivamente a lo largo de este último año y medio a medida que se iban añadiendo artículos a base de datos. Ya hemos introducido más de la mitad del volumen IV, esto es, todos los lemas entre βασιλευτός y el final de la letra beta y una serie de lemas de la letra gamma. La base de datos nos permite conocer algunas curiosidades estadísticas: en los 3906 lemas introducidos correspondientes a la letra beta, son citados 872 autores de los aproximadamente 2000 censados en las listas del diccionario. Actualmente un 80 por 100 de las citas hasta ahora introducidas pertenecen a un grupo de unos 500 autores y los nombres se repiten la mayor parte del tiempo. Estos 872 autores comprenden 1303 obras y totalizan 17106 citas. El autor individual más veces citado es Hesiquio, con 568 citas, seguido de Homero con 526 citas, tras él Plutarco, con 419 citas y a continuación Esteban de Bizancio, con 395 citas, e Hipócrates, con 388 citas. Ello demuestra que hay mucho léxico médico en las palabras entre βασιλευτός y el final de la letra beta, porque también Galeno se encuentra entre los veinte autores más citados. La presencia en los primeros puestos de la lista de autores como Esteban de Bizancio o Ptolomeo es una clara muestra de la abundancia de nombres propios en la letra beta. Hemos hecho que, para el ordenador, «Papiros» sea convencionalmente un autor, con varias obras, como «Papiros de Oxirrinco», «Papiros de la Sociedad Italiana», etc. «Inscripciones» es otro autor, con obras como *Inscriptiones Graecae*, *Supplementum Epigraphicum Graecum*, etc. Esto nos permite recuperar todos los papiros o inscripciones citados en el diccionario conjuntamente. El autor «Papiros» sería el más citado, con 869 citas (esto es, aproximadamente el 5 por 100 de las citas introducidas), seguido de las inscripciones, con 658 citas. El mismo sistema ha sido aplicado a los *Septuaginta* (324 citas), al Nuevo Testamento y a otros textos, como léxicos antiguos o revistas. En el cuadro adjunto puede verse una estadística de los veinte autores más citados en la parte de la letra beta comprendida en el volumen IV del diccionario.

Al final del programa, varios mensajes preguntan si la palabra griega que se está introduciendo está documentada en micénico, si tiene etimología, etc.

La base de datos proporciona la salida de dos tipos de información. En primer lugar, los lemas ya estructurados y con la puntuación adecuada para realizar la revisión definitiva. El paso de información desde la base de datos del diccionario al procesador de textos es directa mediante un procedimiento existente en aquella. El resultado puede ser vis-

EM
Letra BETA
(Autores más citados)



to inmediatamente en la pantalla, pues el programa posee la opción de presentar todas las partes conjuntamente en la forma de una entrada del diccionario. Escribe automáticamente los signos de puntuación, los números y las letras de las divisiones internas, al igual que los distintos tipos de letra (negrita, redonda y cursiva). Es posible convertir la entrada al formato del procesador de textos e imprimirla desde allí en papel.

En segundo lugar proporciona diversa información accesoria como listados de lemas, de autores y obras, de citas, estadísticas de los mismos, etc., es decir, permite realizar un procesamiento de la información según las necesidades del momento posibilitando otro tipo de estudios posteriores.

Su utilidad a corto plazo más interesante es sin duda la de permitirnos realizar con suma facilidad y con mayor fiabilidad y abundancia de información los listados de citas que con gran esfuerzo obtuvimos de la imprenta en los volúmenes anteriores, según explicábamos en la introducción. El siguiente es un breve ejemplo de este tipo de listado, referido a las citas de Eurípides contenidas en una pequeña sección de la letra beta. El listado incluye la siguiente información: 1. Autor. 2. Obra.

3. Referencia. 4. Lema. 5. Apartado del lema en que figura la cita (especialmente útil en el caso de lemas largos). 6. Traducción del apartado en que viene recogida la cita. 7. Información eventualmente asociada a la cita: contexto, traducción del contexto, explicaciones. De este modo resulta muy cómodo el hacer conjuntamente la revisión de las referencias y de los contextos.

E.El.12	βασιλεύω I I	<i>ser rey o reina, reinar</i> χθονός
E.Hec.552	βασιλῆς B I I	<i>mujer del círculo del rey, reina, princesa</i> δούλη κεκλησθαι βασιλῆς οὐσ' αἰσχύνομαι de Polixena
E.Hipp.267	βασιλῆς B I I	<i>mujer del círculo del rey, reina, princesa</i> de Fedra
E.Hipp.1280	βᾶσιληίς I	<i>real</i>
E.IA 1307	βασιλῆς A	<i>perteneciente al rey o a la reina, real, regio</i> εὐναῖσι βασιλίσιν ref. al matrimonio de Hera con Zeus
E.Io 486	βασιλικός I I	<i>propio del rey, real, regio</i> θάλαμοι
E.Io 1087	βασιλεύω I I	<i>ser rey o reina, reinar</i>
E.Med.18	βασιλικός I I	<i>propio del rey, real, regio</i> γάμοι
E.Med.1003	βασιλῆς B I I	<i>mujer del círculo del rey, reina, princesa</i> νύμφη βασιλῆς <i>joven princesa</i>
E.Rh.718	βασιλῆς A	<i>perteneciente al rey o a la reina, real, regio</i> τὰν βασιλίδ' ἐστὶν Ἀτρειδᾶν

Otros listados accesorios con vistas a revisiones de conjunto que la base de datos nos permite hacer son por ejemplo los siguientes: 1. Listado de lemas. Habitualmente se coteja la relación de los lemas de cada fascículo con los de *LSJ* para detectar si falta alguno sin motivo justificado. 2. Listado de los lemas de referencia. En ocasiones sucede que en el artículo al que remite un lema de referencia falta por error la información pertinente. Una revisión sistemática de estas correspondencias permite eliminar este tipo de inconsecuencias. 3. Listado de autores y obras citados. Sobre un listado como éste están hechas las estadísticas que presentábamos antes. Este listado es útil por ejemplo para saber qué autores y obras están desatendidos en el diccionario. 4. Listado de la prosodia. 5. Listado de aparatos morfológicos o de lemas con aparato morfológico. 6. Listado de las referencias al *Diccionario Micénico* (DMic.). Recordemos que el *DGE* está conectado con el DMic., que como se sabe es un anejo al *DGE*, por un sistema de referencias: las propuestas de transcripción al alfabeto griego de las palabras micénicas son incluidas en el *DGE*, remitiéndose a los lemas correspondientes del DMic. 7. Listado de las etimologías.

Todos esos listados no sólo permiten revisar cómodamente de modo

conjunto una misma categoría de información, sino que ayudan a evitar inconsecuencias y a conseguir una mayor uniformidad en la presentación de los artículos.

La base de datos nos permite también generar automáticamente estructuras de artículos, esto es, sin citas, permitiendo una visualización más cómoda de la organización interna del artículo y de los distintos significados de una palabra. Esta información resulta útil tanto como modelo para otros artículos análogos o de la misma familia de palabras como para estudios semánticos o sintácticos. Así por ejemplo, la estructura de dos artículos de mediana extensión como son *βοήθεια* y *βοήθημα* permite advertir la utilidad de esta posibilidad.

βοήθεια, -ας, ἡ I I gener. *ayuda, auxilio, favor, apoyo*; usado como exclamación ¡*socorro!*, ¡*auxilio!*; milit. *ayuda militar, tropas auxiliares*. 2 medic. *ayuda terapéutica, cura, intervención médica*. 3 náut. *cañal de refuerzo* para ceñir el casco de la nave.

II en la administración egipcia *función auxiliar* propia del *βοηθός*, empleado estatal.

βοήθημα, -ατος, τό I *recurso, ayuda*; c. gen. subj.; c. *ἀπό* + gen.; c. indicación de a qué se ayuda; c. *πρός* + ac.; c. indicación de contra qué se ayuda; c. dat. 2 medic. *remedio, tratamiento*; comparando los remedios médicos con los necesarios para acabar una guerra.

Esta última posibilidad abre las puertas a otras sugestivas iniciativas más a largo plazo. En efecto, la base de datos nos permite generar otros diccionarios basados en el *DGE*. Podemos considerar más adelante la creación de un diccionario abreviado para el público universitario, esto es, eliminando determinado tipo de citas (e.g. papiros, inscripciones, autores tardíos, etc.). Esta posibilidad nos permitiría incluso acometer la realización de diccionarios especiales, por géneros, fechas o tipos de textos. Podríamos también acometer la realización de un diccionario escolar, sin citas en absoluto y eliminando las traducciones referidas a determinado tipo de autores. La comparación entre un artículo del *DGE* y lo que serían los artículos correspondientes en uno y otro tipo de diccionario puede ayudar a comprender mejor esta posibilidad.

Artículo completo:

βᾶσιλης [*βασιλῆς* Balbill.30.3] I adj. fem. *real τιμή* II.6.193, Hes.Th.462, E.Hipp.1280, Isyll.64, Orph.Fr. 86A, Th.11.13, Orác. en D.S.35.13, *ἀρχά* Ariphro 1.4, *Epigr.Gr.* 1027.10 (II-III d. C.), *κούρη* A.R.3.886, *μορφή* MAU 1(9).343 (IV d. C.), de Roma, *GVI* 1321.5 (Biti-

nia II d. C.), *TAM* 5(1).208.4, Orác. en *Theos.Tub.*22. 2 subst. ἡ β. *reina* μόρφα βασιλήϊδος ἄμμας Balbill.I.c.

Artículo resumido:

βᾱσιλῆϊς adj. fem. *real* τιμή II.6.193, Hes. *Th.*462, *E.Hipp.*1280, *Orph.Fr.*86A, *Th.*11.13.

Artículo resumido y sin citas:

βᾱσιλῆϊς adj. fem. *real*.

Otras posibilidades a medio o largo plazo son la creación de un diccionario español-griego y de un índice inverso de los lemas (ordenados empezando por el final). En el primer caso, se trataría de recoger por orden alfabético las traducciones dadas en el *DGE*, remitiendo a los lemas y apartados de los mismos en que figuran dichas traducciones. Podría pensarse incluso en un diccionario de este tipo que incluyese citas de autores antiguos documentando las palabras griegas traducidas por tal palabra española. Se podría también agrupar las traducciones identificadas por epígrafes como medic., náut., bot., etc., para tener una relación de vocabulario técnico médico, náutico, botánico, etc., recogido en el diccionario. En cuanto al índice inverso podría considerarse la posibilidad, a la espera de publicar el índice completo, una vez finalizado el diccionario, de que cada fascículo viniese acompañado de un listado inverso de los lemas en él incluidos, que sirviese de complemento a los índices inversos existentes, basados en el *LSJ*⁷.

III.7. *Revisión del original*

En el apartado anterior ya hemos explicado las grandes ventajas que para la revisión de las citas y de otro tipo de información posee la base de datos del diccionario. Podemos mencionar aquí las ventajas añadidas que para esta tarea ofrece la consulta del CDROM del *TLG*. El software que venimos utilizando permite acceder a cualquier autor del CDROM sin necesidad de haber hecho antes una búsqueda. Para muchos autores, por tanto, cuya edición coincide con la nuestra, se puede utilizar esta posibilidad para comprobar citas rápidamente y sin necesidad de acudir al libro, además de en la fase de redacción, en el momento de corregir pruebas o de revisar sistemáticamente las citas. Si la cita

⁷ Cf. especialmente P. Kreschmer y E. Locker, *Rückläufiges Wörterbuch der griechischen Sprache*, Gotinga 1965.

contiene algún tipo de error siempre cabe la posibilidad de efectuar una búsqueda.

III.8. *Imprenta*

La importación de un lema desde la base de datos al procesador de textos ya incorpora, como decíamos más arriba, toda la información tipográfica que será utilizada en la imprenta, de modo que normalmente sirve para realizar una revisión de conjunto antes de ser enviada a aquélla. En este punto, el original revisado puede transformarse, mediante un pequeño programa creado a tal efecto, al formato de imprenta, que mediante un sistema de códigos permite la impresión directa de los lemas con el formato y tipos adecuados. El envío a la imprenta del original del diccionario, ya revisado, en formato de ordenador supone un ahorro de tiempo y trabajo incalculable. Tradicionalmente, la revisión de las sucesivas pruebas de imprenta, nunca totalmente fiables, suponía para los miembros del equipo un gasto en tiempo y esfuerzo superior probablemente a la propia redacción de cada volumen. Actualmente, la imprenta únicamente tiene que hacer una conversión automática de la información contenida en el disquete que le enviamos, en lugar de componer el libro. De este modo, los errores que corregimos en las pruebas de imprenta son achacables, salvo excepciones, al original enviado y no a la propia imprenta.

IV. PROYECTOS

Entre los proyectos inmediatos, se encuentra en primer lugar la ampliación y mejora del equipo. Dada la gran utilización de los ordenadores, se considera necesaria la ampliación de la red con nuevos puestos de trabajo. También está prevista la mejora de los equipos actuales, dotándolos de mayor potencia de trabajo. Por otra parte, la información contenida en la base de datos y otro tipo de información puede ser almacenada en dispositivos ópticos WORM (Write Once, Read Many) de forma que se tenga un acceso a ella de forma similar a como se accede al CDROM. También está en estudio el desarrollo de un procedimiento de conversión automática de la información contenida en el procesador de textos a la base de datos. Por último, entre los proyectos más a medio o largo plazo, cabe mencionar la puesta a punto de un sistema de autoedición del diccionario. Disponiendo de un sistema de autoedición

avanzado de alta resolución puede iniciarse la elaboración de los próximos volúmenes del *DGE*, con un ahorro considerable en los costes de impresión. En este apartado habría que incluir también la utilización de un scanner y de un programa de reconocimiento óptico de caracteres (OCR) para iniciar la conversión a soporte informático de los volúmenes ya publicados del *DGE*, con vistas a futuras reediciones corregidas y aumentadas.

JUAN RODRÍGUEZ SOMOLINOS e IGNACIO ÁLVAREZ