

## El banco de datos del *TLG*, el *DGE* y la lexicografía griega<sup>1</sup>

Francisco R. Adrados  
Real Academia Española

Juan Rodríguez Somolinos  
Instituto de Filología, CSIC

### I. Introducción

El banco de datos del *TLG* se ha convertido ya en un instrumento de trabajo de gran importancia en las diversas disciplinas que tienen por objeto el estudio de la antigüedad y por punto de partida los textos griegos y la lengua griega antigua. Algunas de sus aplicaciones eran ya sabidas o presumidas hace años, cuando el proyecto apenas si había dado sus primeros frutos. Otras se han ido descubriendo con posterioridad, a partir del momento en que los textos empezaron a distribuirse entre la comunidad investigadora, primero en cintas magnéticas y luego en soporte CDROM.

Con ayuda de los programas adecuados, los estudiosos no solamente utilizan (en ocasiones sin admitirlo) el *TLG* para localizar pasajes, buscar paralelos para alguna expresión, identificar papiros literarios o crear un *corpus* de textos como base de algún trabajo de tipo filológico, literario o lingüístico. La bibliografía científica que utiliza el *TLG* abarca también campos tan

---

1. Trabajo leído en la Conferencia Internacional *Thesaurus Linguae Graecae et Thesaurus Linguae Latinae: New Directions in Greek and Latin Lexicography*, Irvine 17-18 de diciembre de 1993. Damos las gracias a los profesores Theodore Brunner and William Johnson, miembros del equipo del *TLG*, y al Patrick Sinclair, profesor en el Departamento de Clásicas de la UCI, por su invitación y por el interés puesto en nuestro proyecto. Otra bibliografía sobre esta cuestión por miembros del equipo del *DGE* incluye los siguientes artículos: J. Rodríguez Somolinos - I. Álvarez, «Informática y lexicografía: la experiencia del *Diccionario Griego-Español*», *Emerita* 59, 1991, pp. 81-99 [= n° 15]; J. Rodríguez Somolinos, «La lexicografía griega en los últimos años», *Eclás* 33, 1991, pp. 83-118; F.R. Adrados, - J. Rodríguez Somolinos, «The *Diccionario Griego-Español* and Byzantine Lexicography», *JÖB* 42, 1992, pp. 1-11 (conferencia leída en la Mesa redonda *Thesaurus et lexica linguae graecae medievalis del XVII. Internationalen Byzantinistenkongress*, Moscú 1991) [= n° 11]; E. Gangutia, «El volumen III del *Diccionario Griego-Español*. Presentación y previsiones futuras», *Actas del VIII Congreso Español de Estudios Clásicos (Madrid, 23-28 de septiembre de 1991)*, Vol. I, Madrid 1994, pp. 109-113 [= n° 1].

espinosos como la crítica textual, los estudios estadísticos para dilucidar cuestiones de autenticidad o cronología de obras literarias, los estudios de morfología y sintaxis con propósitos de investigación o pedagógicos, etc.

Pero la aplicación que a menudo se cita como más evidente es la lexicografía, en sus diversas facetas. A pesar de ello no son demasiados, por el momento, los trabajos puramente lexicográficos o los proyectos lexicográficos de cierta envergadura que se reconocen deudores del banco de datos del *TLG*. Nosotros somos uno de ellos (otro es por ejemplo el *Lexicon zur byzantinischen Gräzität*, que dirige el profesor Trapp en Bonn) y estamos aquí para hablar de ello con la experiencia adquirida en los últimos años. Uno de los fabricantes de *software* para acceder al *TLG* nos dijo en cierta ocasión que nosotros éramos su más «heavy user», pues le hacíamos observaciones a su programa derivadas de un uso exhaustivo del CDROM C.

## II. Nuestra experiencia

### 1. Antes del CDROM

Nuestra experiencia arranca de la época en que los textos del banco de datos del *TLG* se distribuían en cintas magnéticas, un soporte ciertamente incómodo y que requiere para su procesamiento grandes equipos con los que no contábamos y para el cual no existían programas de ningún tipo. A pesar de ello, iniciamos una línea de trabajo que consistía en hacer índices de palabras de baja frecuencia en autores importantes tradicionalmente desatendidos en los diccionarios. Para ello contamos con la colaboración del Centro de Cálculo de la Universidad Complutense de Madrid. No sin esfuerzo logramos crear índices de las formas que aparecen hasta tres veces en Galeno y Pseudo Galeno. Estos índices fueron despojados como un libro más para nuestro diccionario.

### 2. CDROM C

#### 2.1 Consideraciones previas

La incorporación del CDROM a las tareas preparatorias de nuestro diccionario ha sido progresiva. Primero hubo que estudiarlo, tener una idea clara de su contenido, conseguir y adaptar el *software*, bastante experimental e imperfecto al principio. Hubo que valorar los posibles modos de trabajar con él en función de la situación del original y de las posibilidades que ofrecía el *software* disponible. También hubo que acostumbrar a parte del equipo a este nuevo modo de encarar la recogida de información lexicográfica. Hoy por hoy su utilización en nuestro proyecto es cada vez más sistemática, pero todavía no es la óptima.

El aprovechamiento para el *DGE* del material contenido en el CDROM C del *TLG* plantea un problema básico. 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*. Los criterios para seleccionar las ediciones, aunque similares, no son los mismos. Por otra parte, la manera de catalogar los autores y las obras anónimas, ya se trate de la misma edición o de una distinta, también difiere en bastantes casos. De otro lado, el *TLG* incluye en su elenco de autores y obras un buen número de entradas que nosotros hemos desechado, como es el caso no solo 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. 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. Evidentemente, los propósitos de una y otra lista son distintos.

Por otra parte, los sistemas de referencia empleados difieren en muchos casos, ya se trate de la misma edición o de otra distinta. Con respecto a esto, ha resultado ser de gran utilidad el apéndice que acompaña a la tercera edición del *Canon* en que consta el sistema de referencia utilizado para cada una de las obras del banco de datos. En cualquier caso, el *Canon* del *TLG* se ha convertido en un utilísimo instrumento de trabajo para nosotros y para la filología griega en general.

Todos estos inconvenientes que hemos ido viendo nos han convencido de la necesidad de disponer de unas concordancias de uso interno entre ambas listas en que consten las diferencias en lo relativo a ediciones, sistemas de referencia, y nuestras asignaciones y abreviaturas para cada obra, etc. Ya habíamos sentado las bases para este trabajo en lo relativo a los textos contenidos en el CDROM C. Ahora habremos de retomarlo para tener en cuenta las novedades del CDROM D.

Aparte de la recogida más o menos sistemática de materiales para incorporar al diccionario (de ello hablamos más adelante), la segunda utilidad del CDROM para la elaboración del diccionario radica en la consulta del mismo, en el proceso de redacción o de revisión del original, para localizar citas difíciles o con errores (por ejemplo citas que provienen de diccionarios o índices antiguos con errores o con sistemas de referencia anticuados (como por ejemplo el *Stephanus* o el índice de *Jenofonte*), e igualmente para consultar textos cuya edición falta en nuestra biblioteca.

Con respecto a la recogida de materiales nuevos, cuando no teníamos medio de acceder al *INDEX* nos planteamos hacer índices o concordancias de determinados autores o grupos de autores que carecían de ellos y también adaptamos el *software* con que contábamos entonces para hacer búsquedas masivas por períodos cronológicos, por géneros literarios o combinando

períodos y géneros. Se trataba de proseguir la línea iniciada cuando los textos eran distribuidos en cintas magnéticas, pero ahora haciendo todo el trabajo nosotros en ordenadores personales.

## 2.2 El INDEX del CDROM C

Esta línea de trabajo perdió buena parte de su interés cuando pudimos disponer de *software* que permitía el acceso al INDEX del CDROM C. Este índice se ha convertido progresivamente en nuestro principal instrumento de trabajo a la hora de encarar la consulta del *TLG*. El programa que hemos venido empleando para acceder al INDEX es *Searcher*, obra de R.M. Smith y otros colegas de la Universidad de Santa Barbara en California. Es un programa ciertamente imperfecto, pero a lo que sabemos el único que permite el acceso al INDEX en el ámbito de los ordenadores compatibles. Algunas de sus principales deficiencias las hemos resuelto nosotros, por ejemplo desarrollando programas de conversión del formato beta a nuestra codificación del griego para poder trabajar con los resultados de las búsquedas directamente en el ordenador e imprimirlos en griego. Para el CDROM D emplearemos un nuevo programa (todavía carece de nombre definitivo) desarrollado por R.M. Smith y D.J. Dumont en entorno Windows. La primera versión de este programa es todavía demasiado experimental como para que su empleo resulte eficaz.

Con respecto al modo de recoger documentación en el INDEX para completar los artículos del diccionario, empezamos dedicando a una sola persona a esta tarea. Ello fue debido especialmente al hecho de que nos encontrábamos en un punto en que el original del cuarto volumen (ya publicado, Madrid, CSIC, 1994) estaba ya en buena medida redactado. Esta persona recogió materiales para la parte del cuarto volumen ya redactada y para aquella en curso de redacción. Este procedimiento se reveló en cierta medida ineficaz pues favorecía la recogida casi en exclusiva de rarezas, en número reducido y a menudo sin interés real.

Posteriormente nos hemos convencido de que nada puede sustituir la consulta del CDROM por el propio redactor en el momento de redactar el artículo, si bien como último eslabón de la cadena, es decir, una vez que el redactor conoce el material que va a introducir en el artículo procedente de otros diccionarios, de otros despojos manuales y otras fuentes. De este modo, su consulta es menos laboriosa y al mismo tiempo se convierte en un poderoso medio de control. Sin embargo, no hay que obviar el hecho de que cuando una misma tarea la realizan paralelamente diez o doce personas, se corre el peligro de que se produzca una cierta disparidad de criterios y se creen desigualdades entre diferentes partes del diccionario. Este es un problema que estamos abordando actualmente para tratar de elevar con cierta uniformidad la calidad del diccionario.

Actualmente, a cada miembro del equipo que se dispone a redactar una sección de diccionario, se le entrega, además de las fotocopias de otros diccionarios que empleamos como punto de partida (*LSJ*, *Stephanus*, *Lampe*, etc.) y además de los materiales procedentes de nuestros propios despojos manuales (a partir de un momento almacenados en una base de datos diseñada al efecto) dos nuevos instrumentos de trabajo.

El primero es una impresión del *INDEX* correspondiente a la sección de diccionario que redacta. En este índice se encuentran, como es sabido, todas las formas de los textos del banco de datos y su frecuencia. La impresión del mismo en un primer momento fue posible gracias a la colaboración de nuestros colegas del Istituto di Linguistica Computazionale de Pisa, y especialmente de A. Bozzi y A. Sapuppo. A partir de un determinado punto, este trabajo lo vamos haciendo nosotros a medida que avanza el diccionario. En el CDROM a cada forma del *INDEX* está asociada la información de en qué autores y obras está documentada. El redactor estudia esa información, selecciona las formas que le interesan y pide al programa que las busque, con la seguridad de que el programa las va a encontrar en un tiempo mínimo. El trabajo con el *INDEX* supone una eficacia infinitamente superior a cualquier otro modo de trabajar.

El segundo instrumento de trabajo que se reparte a los redactores es la sección correspondiente a su parte de diccionario de lo que hemos dado en llamar *ISCAPLIG* (abreviatura en español de Índice Selectivo de los Cien Autores Principales de la Literatura Griega). Este índice, creado por uno de nuestros colaboradores mediante un programa realizado en entorno Hypercard, arranca de la preposición *διὰ* (más o menos la mitad del volumen V del *DGE*) y se va realizando igualmente a medida que avanza el diccionario. Está basado en unas 1100 obras de unos cien autores de los más de setecientos que incluye el CDROM C. Cubre unos doce millones de palabras, esto es, más de la cuarta parte del total de palabras en el CDROM. Está basado en el índice de formas del CDROM y por lo tanto está sin lematizar y no distingue entre nombres propios y comunes. El hecho de que esté sin lematizar por un lado propicia que haya quizá cierto exceso de información sobre una palabra en un mismo autor y también que pueda perderse la pista de algunas formas flexivas (aoristos y perfectos), pero por otro lado facilita el buscar referencias para una determinada forma de la flexión nominal o verbal, cuando ello es de interés.

Es un índice selectivo, no exhaustivo. Recoge las referencias de las formas que aparecen un máximo de seis veces en un mismo autor. Cuando una misma forma aparece más de seis veces en un mismo autor, el índice indica cuantas apariciones hay de esa forma en ese autor y remite al redactor al índice, léxico o concordancia impresa del mismo. En los casos en que el autor carece de este tipo de instrumentos se remite al propio CDROM. Dentro de cada lema los autores están ordenados con arreglo a dos criterios: primero los autores poe-

ticos y después los autores de prosa. Dentro de cada grupo vienen ordenados *grosso modo* cronológicamente.

Este índice está pensado como un medio de control para que no haya omisiones de importancia en los autores importantes de los diversos períodos de la literatura griega (algo más habitual en nuestra tradición lexicográfica de lo que en principio pudiera pensarse). El redactor hace una selección, con la ventaja de que en él consta el número de apariciones de cada forma en cada autor, lo que permite detectar tanto lo raro como lo más frecuente. En la medida de lo posible, las citas vienen dadas con arreglo a nuestras abreviaturas. Cuando la edición del *TLG* es distinta siempre se hace constar sobre qué edición está hecho el índice. El redactor comprueba todas las citas que decide incluir en nuestras ediciones y las adapta a nuestras convenciones (abreviaturas y sistema de referencias), si es que no coinciden en unos u otros.

En cierto modo, el *ISCAPLIG* supone la continuación de la línea de trabajo antes mencionada consistente en crear índices selectivos. Este índice y otros análogos que en el futuro podamos hacer, en función de las lagunas que vayan quedando en la bibliografía especializada (por ejemplo un índice de médicos o de padres de la iglesia) tienen el propósito de adelantar trabajo al redactor antes de consultar directamente el *INDEX* y pretenden que la consulta de éste quede limitada a aquello que no está cubierto mediante otro material de referencia.

### 2.3 Interés lexicográfico del *INDEX*

El interés lexicográfico del *INDEX* es inmenso y su consulta siempre provechosa. No solamente nos permite localizar las rarezas (numerosos hapax y segundas citas, nuevos testimonios de palabras poco documentadas, formas morfológicas de interés) sino, inversamente, nos permite que no falte en el diccionario nada verdaderamente importante, en particular testimonios de autores de primera fila en palabras de frecuencia media o alta. Las palabras raras nos preocupan, pero no estamos obsesionados por ellas. En un principio nos centramos quizá en exceso en este punto. Ahora hemos puesto los medios para facilitar la localización de otro tipo de material, menos excepcional, pero no por ello menos importante. Podríamos decir que los autores raros o los términos raros son los condimentos (y a veces los adornos) del guiso principal, cuyos ingredientes son las palabras de frecuencia media y alta y los autores importantes, es decir, aquellos que deben estar bien recogidos en un diccionario como el nuestro: el *TLG* nos proporciona ingredientes básicos y también condimentos para el gran guiso que es el diccionario.

La experiencia nos enseña que, como era de esperar, los hápax y los términos raros suelen proceder de autores raros, tardíos o de segunda fila, en general autores menos atendidos en la tradición lexicográfica del griego

antiguo. Estos autores se engloban en cuatro grandes grupos, de los cuales proceden la mayoría de las citas de este tipo que extraemos del CDROM: 1. médicos y veterinarios; 2. filósofos neoplatónicos y comentaristas de Aristóteles y Platón; 3. padres de la iglesia; 4. gramáticos y comentaristas de obras antiguas como Eust. y Phot. Junto a estos cuatro grupos, de donde extraemos la mayor parte de las novedades, hay una enorme pléyade de autores en el CDROM que aportan al diccionario materiales lexicográficos de enorme interés. Podemos mencionar a vuela pluma autores como Arist., X., Thphr., Str., Luc., Plu., App., D.H., Arr., cómicos, oradores áticos, etc. Se trata de autores que por la razón que sea no han sido recogidos suficientemente bien en los diccionarios de griego antiguo, a pesar de contar en algunos casos con índices propios. La consulta del CDROM nos permite por tanto abarcar mejor la historia de la palabra, documentándola en las diversas épocas y géneros en que está representada.

Aparte de las rarezas, los sentidos nuevos, los usos nuevos y las construcciones nuevas necesariamente surgen a mayor cantidad de material estudiado o tomado en consideración para su inclusión en el diccionario. La consulta del índice permite además localizar con la mayor facilidad formas morfológicas concretas que interesan *per se* (ejemplos de tal o cual forma casual, de tal o cual número, de tal o cual tiempo verbal, modo o voz, según la especificidad del artículo).

Por último, otra cosa que, en nuestra opinión, definitivamente permite el disco es el evitar repetir las mismas citas que los diccionarios se han ido pasando unos a otros desde el siglo XVII. Podemos seleccionar otras más adecuadas para figurar en un diccionario y la mayor parte de las veces dar un número mayor, aunque el sentido o el empleo sea el mismo. Desde el principio este diccionario ha tenido claro que, en lo que concierne al número de citas, su pretensión era la de doblar en extensión al *LSJ*. En este momento, evidentemente, si quisiéramos, podríamos multiplicarlo a voluntad, pero no se trata de eso en absoluto. Hoy por hoy, nuestra experiencia nos dice que la proporción ideal en número de citas con respecto a *LSJ* podría ser, como media, de uno a tres, esto es, multiplicar el *LSJ* por tres, incluidos papiros e inscripciones.

El INDEX del *TLG* ayuda para aplicar, aunque sea minimamente, criterios estadísticos a la hora de fijar la extensión relativa de los artículos en función del número de citas que documentan cada palabra. Sin embargo, somos conscientes de que esto es algo muy difícil de aplicar rigurosamente. En otros diccionarios que parten de un banco de datos preexistente (como por ejemplo el *Trésor de la Langue Française* de Nancy) ha sido posible aplicar criterios de este tipo con cierto rigor. Por otra parte, existe siempre una tendencia inevitable y en modo alguno condenable a que los artículos pequeños sean proporcionalmente mayores en número de citas a los artículos grandes.

#### 2.4 Algunos problemas del banco de datos

Una vez dicho esto, también hay que decir que la consulta del CDROM, en general y desde nuestro particular punto de vista, no carece de problemas. Algunos de los inconvenientes derivan del *software* existente. Unos programas son mejores o más sofisticados que otros y en general todos pueden ser calificados de experimentales, aunque en sus sucesivas versiones cada vez lo son menos. Hay que evitar criticar al *TLG* por las insuficiencias del *software* existente. Leyendo bibliografía relativa al *TLG*, advertimos que muchos críticos sólo ven la realidad de las cosas a partir del programa informático que usan.

En alguna ocasión hemos criticado el que el CDROM no venga acompañado de ningún programa de búsquedas o de otro tipo. Aunque este hecho siempre entraña complicaciones añadidas, especialmente para los nuevos usuarios, hoy sin embargo somos más moderados en esta crítica, no solamente porque el *TLG* informa puntualmente sobre los programas existentes, sino porque hemos tomado conciencia de dos hechos. Por un lado, esta situación ha estimulado que surjan diversos programas por la iniciativa de diferentes centros de investigación y casas comerciales. Esta competencia ha resultado ser positiva en nuestra opinión. Por otro lado, esto está directamente en relación con algo que nunca agradeceremos lo suficiente al *TLG*, y es el hecho de que no haya distribuido un CDROM hiperprotegido como el del *CETEDOC* o el del *Institut National de la Langue Française*, que vienen acompañados por un *software* propio bastante satisfactorio, pero a cambio de ello, la información del disco es absolutamente imposible de manipular al margen de dicho *software*.

Otros inconvenientes en cambio son intrínsecos y en algunos casos de difícil solución. Ya hemos hablado anteriormente de que algunas ediciones no son las más adecuadas y otras van quedando desfasadas conforme pasa el tiempo. En nuestro caso concreto, la diferencia de ediciones con el *DGE* a menudo se traduce en lecturas divergentes. Este es un problema que conocemos bien y somos conscientes de que actualizar ediciones es el cuento de nunca acabar. En el *DGE* hemos aceptado el reto de, en la medida de nuestras posibilidades, mantener actualizada nuestra lista de autores y citar los textos por la última edición adoptada. El *TLG* y la *APA* deberían considerar este tipo de actualizaciones en algunos casos. Así por ejemplo, el *CETEDOC* asume este hecho como algo natural y a tener en cuenta en las sucesivas versiones de su CDROM.

En la *Newsletter* 10 (de Julio de 1986), el *TLG* comunicaba que estaba estudiando, a instancias del comité sobre el *TLG* de la *APA*, la posibilidad de incluir los aparatos críticos a los textos del banco de datos. Aparentemente, se trata de un proyecto abandonado pues no se ha vuelto a hablar de él. La ausencia de aparatos críticos pensamos que es un problema menor y su inclu-



sión, algo que probablemente entraña diversas complicaciones técnicas, a nuestro modo de ver debe considerarse un objetivo secundario.

Con respecto a los errores de copia y las erratas reproducidas de las ediciones, diremos tan sólo que habitualmente son fáciles de detectar y estadísticamente su número es ínfimo. El problema es que a nosotros, por el tipo de trabajo que hacemos, nos saltan a los ojos y pueden llegar a plantear problemas: en algún caso desgraciado podrían conducir a la introducción de palabras fantasmas en el diccionario. Curiosamente, la mayor parte de las veces los errores proceden de la edición original, lo que por un lado pone de manifiesto el extremo cuidado puesto al copiarla, pero por otro lado en cierto sentido pone en cuestión el modo en que se ha hecho la revisión. Habría sido mejor esforzarse por corregir el mayor número de ellas al revisar el texto copiado o detectándolas en el INDEX y publicar una lista de las mismas, como hacen los autores del CLCLT, quienes califican su CDROM de *opus semper perficiendum*.

También hay que decir que a veces, el número de citas que documentan una forma engaña: podemos encontrar la misma palabra citada por el autor y la fuente que transmite o reproduce ese pasaje, o por dos ediciones distintas del mismo texto (a veces asignadas a autores distintos), o bien se trata de un pasaje de un autor clásico y referencias al mismo pasaje en comentaristas (ahora también escolios), etc. Este problema no es excesivamente grave pero supone un esfuerzo crítico adicional en un trabajo como el nuestro.

Con respecto al INDEX, su principal problema en su estado actual (luego hablaremos de otras posibilidades de mejora más ambiciosas) es que no distingue los nombres propios de los comunes. Esto es un problema grave que habría que haber resuelto en el momento de crearlo.

A estas consideraciones podríamos añadir, lógicamente ya sin ánimo de crítica, que a veces la masa de material que se pone a nuestra disposición puede llegar a ahogar y desconcertar. Todo ello nos lleva a la conclusión de que debemos redoblar el esfuerzo crítico a la hora de recoger información lexicográfica en el Banco de Datos y asimismo de que su consulta debe ser considerada siempre un paso intermedio antes de verificar o completar la información acudiendo a las ediciones y a otra bibliografía impresa. Si en un principio pudimos pensar que en determinados casos podía bastar con la consulta del disco, enseguida nos resultó evidente que esto no es así. En efecto, desde muy pronto entendimos que este banco de datos (al menos por lo que se refiere a la lexicografía pero sin duda también en otro tipo de investigaciones) no es un mero sustituto de los libros. Es otra cosa: es un instrumento de trabajo intermedio con múltiples posibilidades y el usuario consciente siempre tendrá que acudir a los libros en busca de mayor y más precisa información. Y ello por no hablar de los bancos de datos de inscripciones y papiros, textos muy problemáticos y que casi siempre hay que analizar, en

nuestro trabajo como en otros, teniendo en cuenta su historia, sus sucesivas ediciones, sus problemas críticos, la traducción, las notas, etc. Pero esto es algo que no todo el mundo entiende. En una mesa redonda sobre «Epigrafía e Informática» que tuvo lugar con motivo del X Congreso Internacional de Epigrafía Griega y Latina celebrado en Nimes en Octubre del año pasado, tuvimos ocasión de comprobar los celos y la ignorancia que todavía impera en algunos ambientes, especialmente franceses, sobre lo que son en realidad los bancos de datos con textos antiguos.

### III. Posibilidades de futuro

Para terminar, vamos a hablar brevemente de algunas propuestas de futuro para el Banco de Datos del *TLG*, basadas en nuestra experiencia como usuarios del mismo y en nuestro conocimiento de otros bancos de datos, especialmente el del *CETEDOC* y el del *Institut National de la Langue Française*. Ya hemos hablado de la posibilidad de actualizar algunas ediciones y de añadir a los textos su aparato crítico. No insistiremos en ello.

En nuestra opinión las mayores posibilidades de futuro pasan inevitablemente por el desarrollo del *INDEX* mediante procesos parcialmente no automáticos así como por una mejor integración entre el *INDEX* y los textos mediante *software* cuidadosamente diseñado. Nuestra primera propuesta - la que nos parece más obvia - contempla la posibilidad de que el *INDEX* sea un índice completo, esto es, que incluya las referencias. Es sabido que en el *INDEX* consta la información sobre los autores y obras en que figura cada forma, pero no la referencia precisa, que hay que localizar haciendo una búsqueda específica, más o menos rápida según el *software* empleado y según la extensión de la obra en cuestión. Esto supondría que, por así decirlo, todas las búsquedas ya estarían hechas y el acceso a los textos desde el *INDEX* sería muy rápido.

En segundo lugar, habría que plantearse seriamente el acometer la lematización del *INDEX* y paralelamente o con posterioridad el análisis morfológico de las formas que incluye. Aunque tanto la lematización como el análisis morfológico son procesos complejos, ya hay numerosas experiencias en este sentido en Estados Unidos y en Europa. Se puede aprovechar la experiencia acumulada en otros proyectos e incluso su trabajo y sus programas mediante convenios de colaboración. Los diccionarios automáticos de griego antiguo existentes se van alimentando a sí mismos y mucho trabajo ya está andado. Se trataría de pasar el *INDEX* por uno de estos diccionarios y luego hacer una revisión concienzuda.

Estas ideas nos llevan a lo que sería el *desideratum* por excelencia para este banco de datos: que el *INDEX* sea en realidad una gran base de datos en la que para cada forma conste la siguiente información: por un lado (sin

olvidar la identificación de los nombres propios como tales) la referencia completa (autor, obra, ref.), el lema y el análisis morfológico (quizá también la forma inversa) y por otro lado los datos de Canon, asociados a cada autor y obra, relativos a la fecha del autor y la caracterización genérica de la obra.

Partiendo de la premisa de que en el INDEX a cada forma ya estaría asociada toda esta información complementaria, el modo de trabajar consistiría en la posibilidad de hacer diversos tipos de búsquedas en el INDEX (lematizado o sin lematizar) o ir aplicando filtros al INDEX para generar subíndices del tipo que sean a partir de los cuales se reclamarían los contextos y el acceso a la cita en su contexto amplio. En esta hipótesis, las posibilidades que se abren a la investigación mediante el desarrollo del *software* adecuado, son ilimitadas, desde la más sencilla, que sería sacar en cuestión de segundos el contexto de una sola cita a la más compleja, que sería sacar una concordancia lematizada completa de todo el CDROM ordenada internamente de cualquier modo. Esta última posibilidad constituiría el primer paso para un verdadero Tesoro del griego antiguo, si es que alguien siente la necesidad de hacerlo. Posibilidades intermedias serían las de generar de modo casi automático índices o concordancias, globales o parciales, de autores, épocas o géneros con diversas posibilidades de ordenación interna (cronológica, genérica, morfológica), así como índices o concordancias inversos o de nombres propios con las mismas posibilidades.

#### IV. Conclusión

Volviendo al *DGE*, a modo de conclusión diremos tan sólo que, gracias al banco de datos, ahora estamos en situación de elevar la calidad del diccionario con mayor uniformidad, escogiendo las citas con mejores criterios y reduciendo en buena medida la inevitable arbitrariedad y las lagunas más llamativas. Tenemos las mayores esperanzas de que, paralelamente a su nuevo proyecto de informatización del *Thesaurus* latino, nuestros colegas americanos sigan trabajando en el banco de datos griego y perfeccionando esta genial iniciativa, sin duda puntera en su género.

Sin embargo, hemos de decir que, aunque el banco de datos pueda poner a nuestra disposición muchos materiales, más o menos elaborados, ordenados, organizados o matizados, lo que nunca podrá hacer por nosotros, por mucho que se perfeccione, es darnos la interpretación de los textos (con todas las consultas bibliográficas y el esfuerzo de reflexión que ello a menudo conlleva), ni por supuesto la organización de los artículos, que en nuestro diccionario es un asunto central. Creemos que el *DGE* responde a una necesidad diferente, primordial: la actualización de los diccionarios de griego antiguo al servicio de los filólogos clásicos y de los estudiosos de otras materias.