

## OBSERVACIONES SOBRE LA APLICACION DE LA INFORMATICA A LA LEXICOGRAFIA GENERAL CRECOLATINA

José A. OCHOA  
Universidad de León

Tengo que advertir, en primer lugar, que existen sin duda estudiosos mucho más competentes que yo en estas materias, y en consecuencia abordar este tema puede ser osadía por mi parte<sup>1</sup>. En realidad planteo estas pocas páginas casi como una reflexión en voz alta, a modo de consideraciones teóricas que espero no resulten inoportunas, dado que realmente no son la base teórica de ningún proyecto concreto de próxima realización. Trataré simplemente de animar con estas reflexiones la discusión sobre una técnica de trabajo que puede —y quizá debe— modificar el concepto que tenemos de lexicografía. Lo cierto es que el concepto y uso del diccionario está en progresivo cambio<sup>2</sup> y que en nuestro país se suscita cada vez con más interés este problema, como muestra la aparición del artículo «Informática y lexicografía: la experiencia del *Diccionario Griego-Español*»<sup>3</sup>.

La naturaleza de las lenguas griega y latina y su transmisión les confieren unas características particulares que hay que tener muy en cuenta en la confección de diccionarios bilingües latín/griego - lengua moderna: existe un corpus cerrado que sirve de fuente de información exclusiva, se trata de lenguas de transmisión literaria y por tanto con formas que no sufren la variación constante de significado que las modernas habladas, y por esto mismo tiene especial importancia el apoyo de las definiciones y traducciones con citas.

Podríamos distinguir dos formas de uso de la informática en lexicografía: por un lado, la de auxiliar, que sirve para facilitar el trabajo de confección del diccionario y para agilizar labores de control interno del material que se maneja en redacción; por otro lado, la de redacción, para obtener informaciones que no sería posible alcanzar sin contar con la asistencia de datos memorizados. Tal es el caso de uso del *TLG*, base de datos en proceso de confección que contiene ya buena parte de la literatura griega

---

<sup>1</sup> Existen muchos equipos de trabajo e incluso centros de investigación que dedican buena parte de su actividad a estos menesteres. Citaré tan sólo el ejemplo del *Istituto di Linguistica Computazionale* del C.N.R., en Pisa.

<sup>2</sup> Cf. KAY, M. «The Dictionary of the Future and the Future of the Dictionary», *Linguistica Computazionale* 3 (1983) 161-174.

<sup>3</sup> J.R. SOMOLINOS e I. ALVAREZ, en *Emerita* 59 (1991) 81-99.

desde Homero al siglo VI d.C.<sup>3 bis</sup>

A su vez existen dos niveles de trabajo para cada uno de los objetivos, que en algunos casos condicionarán el método: la producción en papel de los resultados obtenidos, y la producción y distribución en soporte magnético, para su consulta y manipulación mecanizada por parte del usuario.

Ante estas consideraciones, se nos plantea el problema de definir qué es un diccionario informatizado. A este respecto, A. Zampolli estableció una clasificación del material léxico<sup>4</sup>, en el que distinguía los siguientes tipos:

a) «Texts in MRF»<sup>5</sup>, conjuntos lo más amplios posible de textos memorizados que se ponen a disposición de los investigadores, que si son registrados con criterios homogéneos o con una finalidad específica son llamados «Textual Data Bank»<sup>6</sup>. Cuando sobre éstos se puede realizar una consulta interactiva y análisis a través de programas adecuados, se pueden denominar «Textual Data Base»<sup>7</sup>. A partir de estos últimos, los «Lexicographical Textual Data Bases» constituirían la base documental para la edición de un diccionario, con una colección de citas lematizadas extraídas de esa base de datos textual<sup>8</sup>.

b) «Machine Dictionaries». Las «Bases de données dictionnaires» son conjuntos de elementos seleccionados y organizados con ayuda de ordenador para construcción y estructuración del material básico de lemas de diccionario<sup>9</sup>. «Lexicographical Data Base», en cambio, es una estructura semejante a la de una base de datos, en la que el redactor introduce las

---

<sup>3 bis</sup> Sobre su aprovechamiento léase OCHOA, J. A. «Informatic Support to General Greek Lexicography», *Revue* 27 (1991) 161-184.

<sup>4</sup> «Perspectives for an Italian Multifunctional Lexical Database», *Linguistica Computazionale* 4-5 (1987) 301-341.

<sup>5</sup> MRF = Machine Readable Form (formato legible por el ordenador).

<sup>6</sup> Tal sería el caso del *Thesaurus Linguae Graecae* (TLG), de Irvine (California).

<sup>7</sup> En esto se ha convertido el TLG cuando a él se aplican programas como DCU de Ibycus, Pandora o SNS-Greek. En general véase J.A. OCHOA, *SNS-Greek. Manuale Utente*, Pisa 1989.

<sup>8</sup> Sobre la lematización de textos latinos ver BOZZI, A. «Sistema per la redazione semiautomatica delle voci», en FATTORI, M. y BIANCHI, M. *Spiritus IV<sup>o</sup> Colloquio Internazionale*, Roma 1983, p. 567-577, y BOZZI, A. y CAPELLI, G. «The Latin Lexical Database and Problems of Standardization in the analysis of Latin Texts», en *Datennetze für die historischen Wissenschaften?*, Leykam 1987.

<sup>9</sup> QUEMADA, B. «Bases de données informatisées et dictionnaires», *Lexique* 2 (1982) 101-120. AMSLER, R.A. «Machine-readable Dictionaries», *Annual Review of Information Science and Technology*, 19 (1984) 161-209. Para la lengua latina ver LOMANTO, V. «Lessici Latini e Lessicografia automatica», *Memorie dell'Accademia delle Scienze di Torino* 4 (1980).

informaciones para después generar automáticamente el texto de un diccionario en formato utilizable para fotocomposición<sup>10</sup>. Los «Printed dictionaries in machine readable form» corresponden a la operación contraria, la que consiste en memorizar los datos de un diccionario impreso, a través de diversas técnicas (lector óptico, a partir del texto de fotocomposición o tecleando directamente la información)<sup>11</sup>.

Además se puede distinguir entre «Dictionaries in MRF for computer program use» con información codificada para el tratamiento del lenguaje natural (lematizadores, analizadores, traductores bilingües), y los que se usan para investigación lingüística y psicolingüística, con descripciones más o menos formalizadas de las propiedades de las unidades léxicas<sup>12</sup>. Un último tipo serían los «Multifunctional Lexical Data Bases», consistentes en una colección de unidades léxicas en MRF, estructuradas con métodos de base de datos, para uso directo de la información según una multiplicidad de criterios a tenor de los intereses del usuario<sup>13</sup>.

c) «Integrated Linguistic Data Bases», sería el paso siguiente, un conjunto estructurado de bases de datos interrelacionadas de diferente naturaleza (textual, léxica, gramatical, bibliográfica, sociolingüística, etc.) asociadas a programas de análisis lingüístico computerizado para consulta de diccionarios, investigación cuantitativa y estadística<sup>14</sup>, acceso a diferentes datos, y todo ello inserto en una organización que proporciona mantenimiento, distribución, actualización, grupos de trabajo especializado, intercambios internacionales, orientación y enseñanza al usuario, y un

---

<sup>10</sup> MADSEN, H. «Compulexis. A universal dictionary system», en *Abstracts of ZURILEX 86*, Zurich 1986.

<sup>11</sup> BENBOW, T. y WEINER, E. «Machine-readable Dictionaries for the General Public», en D. WALKER y otros [eds.], *Automating the Lexicon: Research and Practice in a Multilingual Environment*, 1987. ENGEL, G. y MADSEN, B.N. «From dictionary to database», en HARTMANN, R.R.K. [ed.], *LEXETER '83 Proceedings*, Tubinga 1984, p. 339-343.

<sup>12</sup> SCHREUDER, R. «Using lexical databases in Psycholinguistics Research», en WALKER [ed.], *Citado*. OLNEY, J. Y RAMSEY, D. «From machine-readable dictionaries to a lexicon tester: Progress, plans, and an offer», *Computer studies in the humanities and verbal behaviour* 3 (1972) 213-220.

<sup>13</sup> ZIMMERMANN, H. H. «Multifunctional Dictionaries», *Linguistica Computazionale* 3 (1983) 279-288.

<sup>14</sup> Ensayo de aprovechamiento de estudio lexicométrico con textos griegos es el ejemplo de MACTOUX, M.M. «Accusés et différenciation lexicale. A propos de deux plaidoyers de Lysias», *Revue* 21 (1985) 139-153. En general ver ALKSEEV, P.M. *Statistische Lexicographie*, Bochum 1984.

largo etcétera. En este etcétera, yo incluiría la posibilidad de aplicar sistemas expertos que con toda esa masa informativa estructurada fueran capaces de proponer soluciones a problemas de análisis fonéticos y morfológicos, definición de etimologías o asociación semántica de unidades léxicas.

No todos estos tipos de material en soporte magnético tienen una utilidad inmediata en la aplicación a lenguas antiguas de larga transmisión especialmente literarias, como el griego y el latín. Creo que podría ser adecuado plantearse una serie de peculiaridades de lo que podría ser una «Base de Datos Léxica Multifuncional». En un diccionario semejante, la estructura de cada lema no tendría por qué dejar de ser fiable desde la ciencia semántica, sino que debería adquirir otra organización: frente al reparto del material analizado en función de la estructura interna del lema, en términos relativos; su distribución en función de una estructura general del vocabulario, en términos absolutos. No sería imposible hallar el punto de encuentro de estas dos formas de concebir la información léxica, que no habría de dejar de realizarse con los criterios científicos con que se viene haciendo.

Aprovechando los sistemas de abreviaturas de los diccionarios, previamente agrupados con un criterio temático, los campos se estructurarían de este modo:

**gramática (datos de caracterización morfosintáctica)<sup>15</sup>**

parte de la oración (N, V, adj., Adv.)

flexión

género

**régimen (datos de rección en la complementación del verbo)<sup>16</sup>**

ac. = acusativo

gen. = genitivo

dat. = dativo

con o sin prep.

---

<sup>15</sup> Sobre la clasificación de los datos gramaticales en un diccionario inglés por ordenador ver la fig. 1 de NOEL, J. «L'informatique lexicographique», *Revue* 19 (1983) 147-174. Sobre analizadores morfológicos del griego ver el estudio de DAVISON, M. E. «Computer analysis of verb forms in the Greek New Testament», *ALLC Bulletin* 11.3 (1983) 68-72.

<sup>16</sup> Conviene tener en cuenta las consideraciones de DELATTE, L., GOVAERTS, S. y DENOZ, J. «Quelques observations statistiques relatives à la subordination en latin», *Revue* (1982).

**dialecto** (la información dialectal es de radical importancia para el conocimiento de la lengua griega)

acarn = acarnanio

aqu. = aqueo

arcad. = arcadio

arcad. chipr. = arcadio chipriota

arg. = argivo

beoc. = beocio

cret. = cretense

delf. = défico

dor. = dórico

egin. = egineta

el. = eleo

eol. = eolio

epir. = epirota

etol. = etolio

eub. = eubeo

hom. = homérico

jón = jónico

jón.-át. = jónico-ático

lesb. = lesbio

locr. = locrio

maced. = macedonio

megar. = megarense

mes. = mesenio

mic. = micénico

rod. = rodio

tes. = tesalio

**registro literario** (el caracter literario de la mayoría de los textos conservados hace imprescindible una distinción de tipo genérico en muchos usos léxicos)

com. = cómico

dram. = drama

ép. = épico

lír. = lírico

orat. = oratoria

poét. = poético

trág. = trágico

**registro cronológico** (necesario para reflejar la evolución semántica de una palabra a lo largo del tiempo, o la acuñación o uso exclusivo una cierta época)

arc. = arcaico  
biz. = bizantino  
heleníst. = helenístico  
imper. = imperial (época)  
posthom. = posthomérico  
tard. = tardío

**nivel semántico** (matices en relación con la intención estilística o los niveles impresivo y expresivo del lenguaje, matices de los varios niveles de metáfora y frecuencia de uso)<sup>17</sup>

abst. = abstracto  
aumentat. = aumentativo  
comport. = comportamiento  
conten. = de contenido  
elípt. = elíptico  
esporád. = esporádico  
euf. = eufemístico  
despect. = despectivo  
dim. = diminutivo  
exclam. = exclamativo  
fig. = figurado  
joc. = jocoso  
melior. = meliorativo  
obs. = obsceno  
onomat. = onomatopeya  
paratrág. = paratrágico  
paród. = paródico  
personif. = personificación  
peyor. = peyorativo  
tecn. = tecnicismo  
trasl. = traslaticio

**tipos de nombre** (distinción entre nombre común y los diversos tipos de nombre propio)

antropón. = antropónimo

---

<sup>17</sup> A este respecto hay intentos parciales de sistematización informática: KANNY, A. «A Stylometric study of Aristotle's *Metaphysics*», *ALLC Bulletin* 7.3 (1979) 12-21, USHER, S. y NAJOCK, D. «A Statistical study of authorship in the corpus *lysiacum*», *Computer and the Humanities* 16 (1982) 85-105, y THURY, E. «Analysis of Lucretius' poetic style by computer: methodological considerations and some conclusions», *Revue* (1983) 189-213.

com. = común  
gentil. = gentilicio  
orón. = orónimo  
patron. = patronímico  
teón. = teónimo  
topón. = topónimo

**registro social**  
aticis. = aticismo  
crist. = cristiano  
dial. = dialectal  
erót. = erótico  
fam. = familiar  
pop. = popular  
prof. = profano  
prov. = proverbio  
ultracorr. = ultracorrección

**matiz semántico dependiente de la morfología (ligado a nociones aspectuales entre sintaxis y semántica)**

desiderat. = desiderativo  
distrib. = distributivo  
dur. = durativo  
étn. = étnico  
ingr. = ingresivo  
instr. = instrumental  
intens. = intensivo  
iterat. = iterativo

**ámbito técnico (distintas modalidades de léxico especial, sería un desarrollo de «tecnicismo», en el grupo «nivel semántico»)**

administr. = administración  
agr. = agricultura  
anat. = anatomía  
arquít. = arquitectura  
astrol. = astrología  
astron. = astronomía  
bot. = botánica  
cien. = científico  
cirug. = cirugía  
econ. = economía  
entom. = entomología  
farm. = farmacología

fil. = filosofía  
filol. = filología  
fís. = física  
fisiol. = fisiología  
funerar. = funerario  
gastr. = gastronomía  
geogr. = geografía  
geom. = geometría  
hist. = historia  
ict. = ictiología  
indum. = indumentaria  
jur. = jurídico  
lit. = literatura  
lóg. = lógica  
ling. = lingüística  
mag. = magia  
mar. = marítimo  
merc. = mercantil  
meteor. = meteorología  
mat. = matemáticas  
medic. = medicina  
métr. = métrica  
milit. = militar  
mineral = minereología  
mús. = música  
náut. = náutica  
orn. = ornitología  
polít. = política  
prosod. = prosodia  
psicol. = psicología  
relig. = religión  
ret. = retórica  
teol. = teología  
vet. = veterinaria  
zool. = zoología

Estas son sólo algunas de las posibles clasificaciones, cuya modificación y ampliación vendría con toda seguridad sugerida por una puesta en práctica sistemática. El campo que en un lema dado no tiene contenido, o ni siquiera tiene razón de existir, quedaría vacío, en caso contrario se introduciría en él la información que se considerara precisa, en unos casos recurriendo a simples abreviaturas, en otros incluyendo tanto texto explicativo o ilustrativo como fuese necesario.



Hay que tener en cuenta que esta estructura tiene unas implicaciones más complejas que las que parecen apuntar J. R. Somolinos e I. Alvarez cuando dicen que desde el *DGE* «se podría también agrupar las traducciones identificadas por epígrafes como medic., naút., bot., etc., para tener una relación de vocabulario técnico médico, náutico, botánico, etc., recogido en el diccionario»<sup>18</sup>. La utilidad que esto tendría no dejaría de ser interna para el diccionario, como sistema de control, pero no como criterio científico fiable. De hecho el uso que hace el *DGE* de esas abreviaturas es ocasional y aleatorio, y en modo alguno sistemático. Se emplea en usos técnicos muy particulares y depende del modo de trabajo de cada redactor<sup>19</sup>, de forma que su plasmación como diccionario de usos semánticos independientes sería desigual. No es a estos usos parciales al que me estoy refiriendo, sino a un diccionario estructurado según este tipo de información.

Una estructura como la diseñada para el *DGE* por Gangutia y Fernández<sup>20</sup>, que se califica de formalizada, no es más que la adaptación de un lema de diccionario a una base de datos. Esta, además de un «campo» *palabra-lemma*, contiene otros organizados jerárquicamente que son denominados *área general*, *área específica*, *área traducida*, *área matizada*, que corresponden respectivamente a las clasificaciones de letra mayúscula, número romano, número árabe y letra minúscula (o ausencia de signo) de los lemas del diccionario impreso. A estos se suman otros campos secundarios y el que contiene las citas.

Lo cierto es que cuando se diseña una base de datos hay que tener muy claros cuáles son los fines que se persiguen con ello y cuáles las características de los datos que habrá de acoger. Los llamados «campos» de las bases de datos pueden usarse como meros depósitos de caracteres alfanuméricos, o bien, como contenedores de información homogénea, tipológicamente equivalente, que nos permite recuperar y ordenar datos que tienen relación significativa entre sí.

Una estructura que reproduce la de un lema-tipo de un léxico — por muy bien pensada que esté para comprender toda la casuística de la organización semántica en una entrada de diccionario — no responde a lo que es una base de datos. A efectos de clasificaciones no sirve aquella información que sea ocasional, sino aquella que resulte uniforme.

Me explico con ejemplos: en el programa diseñado para el *DGE*, en

<sup>18</sup> En el art. cit. en n. *Emerita* 59 (1991) 97 (en el apartado III. 6).

<sup>19</sup> Aunque se trata de un sistema colectivo de trabajo, se pueden apreciar ciertas variaciones en la estructura de unos artículos a otros.

<sup>20</sup> Según se indica en el artículo de Somolinos y Alvarez, III. 6, p. 91.

la llamada área general —lo que corresponde a «A» en el lema— podremos encontrar datos como los que siguen<sup>21</sup>:

lema	área general (A, B, C...)
<i>ἀνακρίνω</i>	A <i>interrogar</i> B <i>diferir, posponer</i>
<i>ἀμεινιός</i>	A I 1 ... B <i>no permanente</i>
<i>ἀλλήλους</i>	A <i>usos recíprocos</i> B <i>usos no recíprocos</i>
<i>ἀμήχανος</i>	A <i>de seres animados</i> B <i>de seres inanimados</i> C <i>adv. ως</i>
<i>ἀπαντάω</i>	A <i>c. suj. de pers.</i> B <i>c. suj. de cosa o abstr.</i>
<i>ἀλογος</i>	A <i>ref. a la proporción y al número</i> B <i>c. ref. a dar cuenta o explicación</i> C <i>ref. al lenguaje</i> D <i>adv. ως</i>
<i>ἀπάγω</i>	A <i>en cont. de alejamiento</i> B <i>en cont. de acercamiento</i> C <i>indif. a la dirección con respecto al que habla</i> D <i>imper.</i>
<i>ἀμαυρόω</i>	A <i>v. act.</i> B <i>v. med.-pas.</i>
<i>ἀμελέω</i>	A <i>frec. en lýtotes</i> B <i>imperat.</i>

<sup>21</sup> Extraigo los datos del fascículo II del *DGE*, Madrid 1986.

ἄλς	A	ή ἄλς
	B	ὁ ἄλς
	C	ἄλα
ἀμφότερος	A	I ...
	B	sg. y neutr. sg. o plu., nunca en época arcaica salvo en forma adverbial
ἀτίτυπος	A	adj.
	B	sust.
	C	adv.
ἅμα	A	adv.
	B	prep.
	C	conj.
ἄνθρωπος	A	I ...
	B	B medic. ὁ ἄ. el cuerpo humano

En este campo, por tanto podremos encontrar traducciones al castellano (*ἀνακρίνω*), si bien en algunos casos puede quedar la casilla vacía, porque se define en los niveles inferiores (área específica, traducida..., como en *ἀμενηνός*, *ἀμφότερος* y *ἄνθρωπος*). También puede haber rasgos semánticos definidos y no traducidos (*ἀλλήλους*), en algunos casos contextuales como el tipo de sustantivo con el que va un adjetivo (*ἀμήχανος*), o el sujeto con el que va un verbo (*ἀπαντάω*), o el sentido general de contexto (*ἀπάγω*); a veces combinado con categorías gramaticales (*ἄλογος*). Podemos hallar división por categorías gramaticales como la voz (*ἀμανρόω*)<sup>22</sup>, el género (*ἄλς*) o el número (*ἀμφότερος*), o bien la combinación de una figura estilística con un modo verbal (*ἀμελέω*), y también por clases de palabras (*ἀντίτυπος* y *ἅμα*). A este nivel podemos encontrar incluso una clasificación por uso técnico de la palabra (*ἄνθρωπος*).

Conste que no estoy valorando los criterios semánticos aplicados a estos ejemplos<sup>23</sup>, sino su utilización como base de la estructura de un banco

<sup>22</sup> Tampoco debemos esperar que estos criterios se mantengan siempre en este nivel de área general, porque para el lema *αμβλύω*, por ejemplo, la categoría «voz» está asignada al área específica.

<sup>23</sup> Criterios que, por otra parte, son muy válidos y de gran interés, con una aplicación experimental de la semántica estructural. Para los criterios científicos del

de datos. Lo cierto es que lo único que esos datos —incluidos en el mismo campo— tienen en común es el haber sido colocados por el redactor en una posición jerárquicamente superior en la organización de la semántica de una palabra. ¿Qué interés tendrá entonces ese tipo de campo en una base de datos léxicos?

Si seguimos las indicaciones de Somolinos y Alvarez<sup>24</sup>, la primera utilidad es la de convertir la información en un texto continuo, pasando de la base de datos al procesador de textos, para lo cual no necesitábamos partir de una estructura que además hace muy incómoda (si no casi imposible) la redacción en pantalla. El segundo tipo de empleo está siempre orientado a la creación de listados de uso interno para control del propio diccionario. Un tercer tipo sería el de crear diccionarios impresos con sólo algunos datos seleccionados de entre todo el material, como un diccionario escolar, para lo cual habría que esperar a ver terminada la *editio maior* del DGE. De todas formas, siempre correremos el riesgo de ver reducidos los lemas *ad absurdum*, por la automatización de un proceso que por el contrario necesita criterios de selección distintos. Una reducción, como la del ejemplo propuesto por estos autores, ocultaría aspectos interesantes para un objetivo propedéutico<sup>25</sup>, como el régimen y la construcción habitual de los complementos de un verbo<sup>26</sup>, o ejemplos que ilustran a qué tipo de sustantivo suele ir aplicado un adjetivo<sup>27</sup>, o la sustantivación de un adjetivo o participio<sup>28</sup>. Incluso se omitirían ciertos usos particulares o acepciones, como ocurre con *Βασιλεύς* que sin artículo con frecuencia es «el rey Persa», lo que interesará mucho en un diccionario manual, mientras que en el DGE, queda lógicamente localizado en una acepción muy particular, en l.1.d. Además, habría que reducir el número de lemas, no sólo su contenido, para confeccionar un diccionario escolar, y esto llevaría consigo el establecimiento de criterios de selección, lo que no sería sencillo ni, en modo alguno, automatizable.

---

DGE puede leerse E. GANGUTIA [ed.], *Introducción a la lexicografía griega*, Madrid 1977, y F. R. ADRADOS, «The Greek-Spanish Dictionary and Lexicographic Science», *Lexicographica* 2 (1986) 8-32, además de otros trabajos de este autor en los volúmenes: *Estudios de semántica y sintaxis*, Barcelona 1975 y *Nuevos estudios de lingüística Indoeuropea*, Madrid 1989.

<sup>24</sup> Cf. art. cit., p. 93-94.

<sup>25</sup> En este caso los ejemplos los tomaremos del fascículo III del DGE, Madrid 1991.

<sup>26</sup> El uso de *ἀπειρίζω* con acusativo, con dativo o con sintagmas preposicionales (*πρός, εἰς*) aparece reflejado en el DGE sólo en los ejemplos de las acepciones.

<sup>27</sup> *ἄρρηκτος*, 2 «inquebrantable, indestructible», aplicado a *ἀρεταί*, pero también a *πόλις*.

<sup>28</sup> La sustantivación *τὸ ἄφωνον* «consonante», que está dentro de la acepción *ἄφωνος* II. 2, desaparecía en la reducción.

Me parece muy delicado que se piense que también con este criterio «se pueda acometer la realización de diccionarios especiales, por géneros, fechas o tipos de textos»<sup>29</sup>, porque la lexicografía especial requiere unos presupuestos teóricos y un modo de trabajo que no se puede en modo alguno reducir a un proceso de criba del material de un diccionario general.

Parece, pues, que el usuario del *DGE* no se podrá beneficiar en un futuro de este sistema más que indirectamente, en la medida que permite a sus redactores agilizar su trabajo. A no ser que tenga debilidad por «algunas curiosidades estadísticas», que por otra parte pueden llevar a engaño. Si en casi 4.000 lemas aparecen citados 872 autores<sup>30</sup>, esto no quiere decir que el resto de los autores conservados no use esas palabras, dado que forzosamente el material del diccionario es selectivo. Si un conjunto de textos como el *Corpus Hippocraticum* aparece citado 388 veces, puede indicar una mayor presencia de léxico médico en ese conjunto de lemas, pero también que se trata de un autor mejor estudiado y del que se han extraído más datos que de otros, y si el 5% de las citas corresponde a textos papiráceos estaremos ante un dato del que será muy difícil obtener una valoración significativa<sup>31</sup>.

En definitiva, retomando las reflexiones iniciales, tendremos que concluir que el destinatario del *DGE* no se beneficiará de esta «informatización», dado que no podrá hacer a su instrumento de trabajo preguntas como ¿qué palabras de la raíz de *δίκη* encontramos en léxico filosófico?, ¿qué sustantivos agentes pertenecen al léxico de la vida privada? ¿qué verbos de movimiento rigen genitivo en la lírica?, ¿qué formas de dialecto dórico se nos atestiguan en lenguaje familiar?

Para que el usuario de un diccionario pudiera interrogarle en este sentido, tendría que estar previamente concebido para ofrecer esta información y sería necesaria una estructura en la que tuviera radical importancia el análisis de los elementos semánticos y morfológicos que integran la palabra-lema. Habría entonces que cambiar la concepción de la información etimológica que habitualmente se usa, convirtiéndola en un verdadero análisis de elementos significativos que integran la palabra (por derivación y composición). Estoy pensando ahora, de nuevo, en un diccionario

---

<sup>29</sup> Cf. art. cit., p. 96.

<sup>30</sup> *Ibidem*, p. 93.

<sup>31</sup> Conviene no perder de vista la bibliografía sobre los problemas metodológicos a este respecto. Sirvan de muestra: TOURNIER, M. «D'où viennent les fréquences de vocabulaire? La lexicométrie et ses modèles», *Mots... Ordinateurs... Textes... Sociétés* 1 (1980) 189-209, y SALEM, A. «Analyse factorielle et lexicométrie. Synthèse de quelques expériences», *Mots... Ordinateurs... Textes... Sociétés* 4 (1982) 148-168.

de griego, donde estos recursos tienen una productividad inmensa. Así, habría que diseñar un campo en el que se incluyera la raíz, y otros en los que se reflejaran los prefijos, sufijos y otras raíces que forman la palabra- lema<sup>32</sup>. De este modo sería posible interrogar al banco de datos sobre los términos técnicos de un ámbito concreto (léxico militar, por ejemplo) de las palabras de una raíz o conjunto de raíces.

Tampoco tendría por qué ser especialmente difícil pasar de una escritura como ésta a una disposición de lema organizado internamente. Bastaría con asignar a cada campo de los utilizados para un lema dado, claves numéricas que permitieran generar la ordenación jerarquizada de ese lema, como habitualmente lo encontramos en los diccionarios.

Pero todo esto, repito, implicaría un planteamiento inicial y unos presupuestos diferentes de los que se establecen para un diccionario tradicional, y obligaría al redactor a realizarse una serie de preguntas al organizar las definiciones y datos que incluye en el lema, porque tendría ante sí una plantilla fija de datos que consignar, y debería decidir si para ese caso concreto tiene o no tiene interés cada casilla de este esquema. Y esto lleva consigo un método particular de trabajo.

Seguro que la hipotética puesta en práctica de semejante proyecto implicaría muchos problemas e imprevistos, pero esta certeza creo que no impide que se pueda discutir acerca de los presupuestos teóricos y los criterios científicos que deberían estar en su base. De cualquier modo, pienso que es útil puntualizar una serie de conceptos y distinguir claramente qué es un diccionario informatizado, para no confundirlo con la gestión interna mecanizada de un diccionario convencional.

---

<sup>32</sup> Aquí tropezamos con dificultades serias de análisis gramatical y etimológico, unas propias del griego y otras de tipo general (algunas de ellas puestas de relieve por RIZAIEV, S. *The Experience of Statistical Investigation of Syllable Structure*, Tashkent 1975).