

SIGNOS, GRAFÍAS Y BANCOS DE DATOS TEXTUALES LATINOS

This paper demonstrates that all the things needed to use Latin texts as data bases is an upper character set (ASCII 128-255 dec.) designed to give full satisfaction to the least graphematic needs for Latin scholars. This paper contributes to settle which are those needs, and to show an easy and efficient way of satisfying them.

Algunos de los que gustan de andar adivinando el futuro, deslumbrados por las posibilidades casi mágicas de la tecnología más reciente, se atreven a profetizar que, no tardando mucho, la imprenta será reemplazada por los *multimedia*, y que todo, desde el periódico hasta la última novela, lo habremos de leer en la pantalla de un ordenador-televisor. Seguramente esas previsiones son exageradas, y en su momento resultará que las publicaciones electrónicas seguirán, como ahora mismo, siendo suplementarias de las impresas, a las que probablemente desplazarán sólo en el sector de las grandes obras de consulta, atlas y enciclopedias.

En el campo de la Filología Latina, puede esperarse que algún día aparezcan ediciones en disco de la *Realencyclopädie* y del *CIL*, y que tengan cierto éxito y cierta continuidad las publicaciones documentales del corte del *incipitaire* que anuncia Brepols¹. Los *lexica* y los diccionarios generales, en cambio, no me parece que hayan de ser carne de *multimedia*, aunque ciertamente son ya los principales beneficiarios de la informatización y, posiblemente, aprovecharán las ventajas que ofrece la tecnología para achicar el tamaño de la publicación impresa minimizando los costes, optimizando la manejabilidad y maximizando el contenido. En efecto, debemos empezar a considerar seriamente la posibilidad de que, en un futuro no muy lejano, haya diccionarios latinos generales y *lexica* que en cada entrada estampen sólo las citas imprescindibles para tipificar el uso o la acepción correspondiente, remitiendo al usuario para

¹ Titulado *In principio*, en un banco de datos, grabado en disco compacto, del *Institut de Recherche et d'Histoire des Textes* de París, con la descripción de aproximadamente cuatrocientos mil manuscritos identificados por su *incipit*.

las demás a un banco de datos complementario, en el que las citas podrán darse no ya *in extenso*, sino *in integro* y hasta *in situ*. Pues, sin entrar en arideces tecnológicas que aquí no vienen al caso, hay que señalar muy especialmente que, con ayuda de un programa de ordenador adecuado, un texto en copia electrónica puede ser manejado como si fuera un banco de datos textuales.

En sus primeros pasos, esa revolucionaria innovación puede tropezar con una dificultad de orden técnico que convendría allanar cuanto antes: los juegos de caracteres previstos actualmente para los ordenadores y las impresoras no satisfacen las demandas del tratamiento mecanizado, en tanto que datos textuales, de las obras literarias, las inscripciones y los papiros que constituyen los fondos documentales de un diccionario latino general. Para remediar esa deficiencia, basta con diseñar una tabla de signos con arreglo a nuestras necesidades en la materia y a las capacidades a este respecto de la tecnología.

En principio, aquéllas son mucho mayores que éstas, puesto que en una tabla que aspire a la sanción de los organismos internacionales de normalización tienen cabida solamente 127 caracteres, número menor que el de los signos alfabéticos del sistema grafemático del griego clásico, al que, en el catálogo de necesidades reales, hay que añadir una cierta cantidad de cifras latinas, un puñado de signos diacríticos, una discreta provisión de signos críticos y convencionales y, en fin, un contingente de grafías modernas, entre ellas nuestra ñ, de las que no se podría prescindir sin quebranto. Es inevitable, pues, una solución de compromiso, que sólo puede alcanzarse mediante la transcripción o la transliteración del griego.

Por supuesto, la transcripción, por lo que tiene de empobrecedora, es un procedimiento absolutamente detestable, y debe ser desechada, lo mismo que las transliteraciones del estilo del conocido «formato *beta*» del *Thesaurus Linguae Graecae*, eficaz pero rudimentario e incómodo. A mi entender, la menos mala de las soluciones técnicamente posibles consiste en una transliteración parcial, reservando para el griego 17 de los 127 puestos disponibles en la tabla de caracteres. Con este expediente, se obtiene en los bancos de datos textuales un griego tolerablemente legible, que luego, en el llamado «proceso de textos», puede ser fácilmente puesto en limpio ².

² La operación puede ser automatizada al pasar del tratamiento de datos a la confección de textos si en el banco de datos se han acotado convenientemente las porciones del texto que deben ir en griego. Por ejemplo, [^]άμπελος[^] → ἄμπελος. El sistema de transliteración que se propone se hallará descrito esquemáticamente al final de esta nota, en la Tabla II. Los diecisiete signos aludidos ocupan en la Tabla I las posiciones B6-B8 (181-184 en expresión decimal) y EO-EC (224-236 decimal).

En cuanto a las cifras latinas, sería en verdad gratificante que todas ellas pudieran figurar en el juego de caracteres, pero eso no es posible y, aunque lo fuera, en un banco de datos textuales serían puramente suntuarias las ventajas de poder expresar con sus símbolos propios la *siliqua*, la *sextula*, el *scripulum* y otras magnitudes fraccionarias. En estos casos, lo más conveniente es, en mi opinión, usar las formas plenas en los bancos de datos, y las representaciones abreviadas en el «proceso de textos»; y lo mismo digo de las *notae sententiarum*, de las *iuridicae* y de las *militares*, así como de las expresiones numerales en monograma³. En cambio, las cifras numerales de uso corriente para expresar los múltiplos del millar y de la centena de millares, el *episemon* y los símbolos del *sestertius*, el *quinarius* y el *denarius*, trece signos en total, deben, a mi juicio, formar parte de la tabla de caracteres normalizada.

En materia de signos diacríticos y ortográficos, a los disponibles actualmente habrá que añadir los correspondientes a los acentos grave y agudo, los de cantidad vocálica breve y larga, el de la *interpunctio* epigráfica, la diéresis y el apóstrofo.

Por lo que respecta a los signos críticos, sabido es que en las ediciones de textos tanto su número como su significado son, en gran medida, discrecionales, dependiendo del criterio de cada editor y de las particularidades de cada texto. En los bancos de datos textuales, que en principio no deben ser copia electrónica fidelísima de las ediciones impresas de las que toman su contenido, conviene adoptar un sistema de signos críticos como el del *Thesaurus Linguae Latinae*, antecedente y en muchos aspectos modelo de los bancos de datos textuales, muy sencillo y ajustado a los usos hoy generalizados, si no regularizados. Para completarlo, bastaría con añadirle dos signos para señalar las lagunas, por omisión o por mutilación, detectadas en el texto: un punteado (...) para aquellas cuya extensión pueda determinarse con alguna certeza, y para las demás un asterisco, que, por razones tecnológicas, debe ser diferente del usado para marcar las formas reconstruidas por conjetura o hipótesis, que junto con los *unci curui*, los *angulati* y los *acuti*, forma parte de la tabla de caracteres principal⁴. En la diseñada especialmente para los bancos de datos textuales

³ Véanse, apuestos a la Tabla I, algunos ejemplos de desarrollo de las cifras numerales y epigráficas. En el banco de datos, la barra que precede a la forma plena indica que ésta aparece cifrada en el texto; en la confección de textos, hace posible la reconversión automatizada a cifras de las formas plenas. El mismo procedimiento se aplica, en el sistema propuesto, a las cifras numerales griegas, como se verá al final de la Tabla II.

⁴ A estas alturas, sólo los más contrarios a interesarse por los detalles técnicos pueden no saber que los ordenadores personales y las impresoras manejan un juego de 256 caracteres, formado por una tabla principal (caracteres 0-127 en notación decimal. 00-7F en hexadecimal), universal, y una tabla secundaria cuya composición puede variar para acomodarla a las particulares necesidades de una lengua, o de las lenguas de un área geográfica

latinos, pues, los signos críticos ocuparán sólo tres plazas: el punteado una, otra el asterisco adicional, y la *crux philologica* la tercera.

El catálogo de los signos precisos para dar una presentación mínimamente satisfactoria a los textos latinos en copia electrónica se completa con un juego de notaciones métricas y prosódicas, cuya configuración debe considerar solamente las necesidades reales que se experimentarán en la compilación de un banco de datos textuales. Éstas, teniendo en cuenta la posibilidad de aprovechar para estos fines otros elementos de la tabla de caracteres, se cifran en solamente siete signos: cinco para representar las vocales largas, uno para la sílaba breve y otro más para la *anceps*.

Esto hace un total de cuarenta y siete signos, quedando en la tabla de caracteres espacio sobrado para los veinticinco caracteres semigráficos de las tablas secundarias hoy homologadas internacionalmente, los treinta y nueve signos alfabéticos, ortográficos y tipográficos que se usan corrientemente en los escritos en las lenguas española, italiana, francesa y alemana, y, finalmente, hasta dieciséis signos más, tres de los cuales conviene reservar para la buena administración de los bancos de datos textuales, en los que se necesitarán para delimitar claramente los ficheros que compongan un banco de datos, las fichas que los formen y los apartados que en éstas se hayan de señalar⁵.

Con trece signos, algunos de ellos usuales en los diccionarios, y todos de significado discrecional, he llenado los huecos que quedaban en la tabla de caracteres que he compuesto para mi uso personal, mientras no adopten una mejor como estándar mis colegas latinistas, a cuya consideración someto este expediente que, a mi entender, satisface las demandas que en materia de signos y grafías plantea el tratamiento de los textos latinos como si fueran bancos de datos textuales.

LOIS C. PÉREZ CASTRO

determinada. Las tablas secundarias homologadas actualmente son las denominadas *PC 437* (U.S.A. y lenguas de la Europa occidental), *PC 850* (*multilingual*), *PC 860* (portugués), *PC 863* (Canadá, francófono), *PC 865* (noruego), *PC 853* y *PC 857* (turco), *PC 437 Greek* y *PC 851* (griego moderno), *PC 855* (alfabeto cirílico), *PC 866* (ruso) y *PC 852* (lenguas de la Europa oriental que utilizan el alfabeto latino con ampliaciones). Estas tablas constan de 128 caracteres (*80-FF hex.*, 128-255 dec.), pero el último de ellos (*FF hex.*, 255 dec.) ha de ser, necesariamente, un espacio en blanco.

⁵ Véase, al final de la nota, la Tabla I.

TABLA I
CARACTERES PARA BANCOS DE DATOS TEXTUALES LATINOS

	B	9	A	B	C	D	E	F
0	Ç	É	Á	⋮	ˆ	ṽ	η	*
1	ü	í	í	⋮	ˆ	ṽ	ή	'
2	é	ó	ó	⋮	ˆ	ṽ	ή	...
3	ā	ō	ú		†	ē	ñ	~
4	ä	ö	ñ	†	-	ō	9	?
5	à	ò	Ñ	θ	†	ñ	ξ	x
6	β	ū	⊖	⊖	ˆ	Γ	χ	÷
7	ç	ù	⊖	ψ	ˆ	Γ	ψ	≈
8	ê	ú	ι	Ω	⊖	ς	ó	°
9	ë	ö	⊖	⊖	ˆ	ˆ	ò	· — final de campo, apartado
A	è	ü	⊖	⊖	ˆ	Γ	ō	· — <u>interpunctio epigráfica</u>
B	ÿ	ā	*	⊖	⊖	ˆ	'	≈
C	î	ē	†	⊖	⊖	ˆ	'	≠
D	ì	ī	ı	ˆ	=	†	"	□ — final de ficha
E	À	ō	«	ˆ	⊖	†	'	■ — final de fichero
F	Á	ū	»	ˆ	ˆ	†	'	

EJEMPLOS DE TRANSCRIPCIÓN DE CIFRAS NUMERALES Y EPIGRÁFICAS ROMANAS

- | | |
|-------------------|-------------------------------------------|
| ε — \semuncia | Ɔ — \Gaia - \centurio_2 - \centuria - \ps |
| ʹ — \sicilicus | 7 — \filia - \femina |
| z — \sextula | 9 — \puella - \pupilla |
| z̄ — \dim_sextula | l — \i_longa |
| ε̄ — \scripulum | γ — \centurio_1 - \centuria |
| ” — \siliqua | ξ — \chrismon |

TABLA II
 TRANSLITERACIÓN DEL GRIEGO CLÁSICO

A α — A a	Ξ ξ — Ξ ξ	ά ά ά — ά ά ά
B β — B b	Ο ο — Ο ο	έ έ έ — έ έ έ
Γ γ — G g	Π π — Π π	ή ή ή — ή ή ή
Δ δ — D d	Ρ ρ — Ρ ρ	ί ί ί — ί ί ί
E ε — E e	Σ σ — Σ σ	ό ό ό — ό ό ό
Z ζ — Z z	ς — ς	ύ ύ ύ — ύ ύ ύ
H η — H η	Ϛ Ϛ — Ϛ Ϛ	ώ ώ ώ — ώ ώ ώ
Θ θ — Θ θ	Τ τ — Τ τ	
I ι — I i	Υ υ — Υ υ	ϊ ι ι — ι ι ι
K κ — K k	Φ φ — Φ φ	ϋ υ υ — υ υ υ
Λ λ — L l	Χ χ — Χ χ	
M μ — M m	Ψ ψ — Ψ ψ	
N ν — N n	Ω ω — Ω ω	

” — ’	ά ά ά — ’a ’á ’à ’â	ά ά ά — ’a ’á ’à ’â
” — ’	έ έ έ — ’e ’é ’è ’ê	έ έ έ — ’e ’é ’è ’ê
” — ’	ή ή ή — ’η ’ή ’ή ’ῆ	ή ή ή — ’η ’ή ’ή ’ῆ
	ί ί ί — ’i ’í ’ì ’î	ί ί ί — ’i ’í ’ì ’î
” — ’	ό ό ό — ’o ’ó ’ò ’ô	ό ό ό — ’o ’ó ’ò ’ô
” — ’	ύ ύ ύ — ’u ’ú ’ù ’û	ύ ύ ύ — ’u ’ú ’ù ’û
” — ’	ώ ώ ώ — ’w ’ώ ’ὠ ’ῶ	ώ ώ ώ — ’w ’ώ ’ὠ ’ῶ

ρ — ’r Ϝ — \digamma ϝ — \koppa Ϟ — \sampi