

## EL HIPERTEXTO ANTE EL RETO DE LOS TEXTOS MEDIEVALES: NUEVAS REFLEXIONES SOBRE INFORMÁTICA HUMANÍSTICA

JOSÉ MANUEL LUCÍA MEGÍAS  
Universidad Complutense de Madrid

### 1. Sobre nombres, conceptos y otros demonios

Hay dos conceptos en el título de esta intervención que quizás merezcan una definición y una explicación; ya que el uso indiscriminado de uno (*hipertexto*) o la escasa utilización del otro (*informática humanística*) pueden situarnos en el nimbo de los conceptos muy utilizados y nunca comprendidos.

1.1. Ted Nelson en su libro *Literary Machines* de 1965 fue el primero en utilizar el término de *hypertext*, aunque con un uso general que se ha ido concretando en los últimos tiempos. En todo caso, en la actualidad se tiende a un uso generalizado del mismo, de tal manera que *hipertexto* se llega a confundir con cualquier identidad informática; y, siguiendo este falso hilo de argumentación, se entiende que una *edición hipertextual* sea aquella que se difunde en soporte informático. Pero no son tan extensas las aguas en las que debe navegar el hipertexto.

Susana Pajares Tosca en su reciente libro *Literatura digital. El paradigma hipertextual* (2004) propone la siguiente definición, que vale la pena reproducir en este momento:

El hipertexto es una estructura de base informática para organizar información que hace posible la conexión electrónica de unidades textuales (de diferente tamaño, categoría y naturaleza) a través de enlaces (links) dentro de un mismo documento o con documentos externos. Requiere la manipulación activa del lector para poder ser leído/utilizado, además de la actividad cognitiva común a cualquier proceso de lectura (2004: 34).

La definición puede parecer en una primera lectura un tanto enrevesada, pero en realidad lo que se está proponiendo es una idea muy simple, de ahí su carácter revolucionario: el *hipertexto* (gracias a la tecnología de codificación informática) consigue transformar el *texto* (junto a la imagen, al sonido) en una *red de información*, compuesta de *nodos* y *enlaces*.

El hipertexto, de este modo, va más allá de la digitalización, es decir del hecho de escanear un texto, una edición, una información para poder colgarla en la Red.

El hipertexto, además, va más allá de la simple acumulación de la información, de esas bibliotecas virtuales o bases de datos textuales que lo único que permiten es la acumulación sin permitir la posibilidad de relacionarla.

El hipertexto, por estas causas, se presenta como una nueva forma de escribir (el *hipertexto creativo*), una nueva forma de difundir el saber (el *hipertexto informativo*) y, por último, en una nueva forma de editar (el *hipertexto textual*).

1.2. La informática humanística es una nueva disciplina científica que analiza y experimenta –de una manera sistemática y, por tanto, científica- las nuevas posibilidades que en el campo de las humanidades se abren gracias al uso de las diferentes herramientas y medios de transmisión creados por las nuevas tecnologías informáticas.

En su formulación románica –más allá de la anglosajona que se mueve en ámbitos bien alejados de la filología- puede decirse que la informática humanística nació en 1990 cuando Tito Orlandi publica su libro *Informatica umanistica* en Italia.

Es en este país donde esta disciplina ha tenido más éxito, estando ya presente en el cursus académico: en enero de 2003 se ha pedido al Ministerio de Educación italiano la creación de una nueva área de conocimiento: *Informática aplicada a las disciplinas humanísticas (informática humanística)*.

Dentro de las diferentes ramas en las que se divide la informática humanística (lingüística, histórica, filosófica, arqueológica...), aquí interesa destacar la *Informática textual*, que es la rama de la Informática humanística que se dedica a estudiar, con la aplicación de las nuevas tecnologías informáticas, todas las cuestiones que rodean al texto, desde su génesis, difusión moderna, análisis y edición. Cuatro son los campos de estudio dentro de la Informática textual a los que deberíamos prestar atención:

- a) *Editorial*: difusión del texto con nuevos soportes (desde la digitalización al diseño del hipertexto), diseño y gestión de ediciones hipertextuales y de la estructura de las bibliotecas virtuales, difusión de hipertextos informativos e, incluso, de hipertextos creativos.
- b) *Documental*: diseño y gestión de archivos hipertextuales y de bases de datos textuales.
- c) *Instrumental*: uso y perfeccionamiento de programas específicos para la edición y análisis de los textos.
- d) *Hermenéutica*: utilización de las herramientas informáticas para analizar los textos desde diversas perspectivas, hasta ahora no exploradas.

La informática humanística –y en concreto, la informática textual– no ha de ser considerado como un sustituto de la filología, sino como una de sus ramas, en donde se utilizan nuevas herramientas que una nueva tecnología ha puesto en nuestras manos en el siglo XXI.

## 2. La informática humanística en el ámbito hispánico: ¿cien años de soledad?

La *Informática Humanística* –o las *Humanities Computing*, para el ámbito anglosajón– se ha ido haciendo un hueco en los dos últimos decenios, en especial en las Universidades y centro de investigación de los Estados Unidos de Norteamérica. ¿Qué sucede en el ámbito hispánico? Y permítanme que me limite a glosar la situación de España, que es la que mejor conozco.

Si en Italia la *Informática Humanística* es ya una asignatura obligatoria en muchas licenciaturas de humanidades, si se está incluso planteando la posibilidad de crear una propia licenciatura y que sea reconocida como área de conocimiento, en España esta disciplina brilla por su ausencia. Sólo es posible descubrir unas pequeñas huellas en el amplio abanico de posibilidades que las universidades españolas ofrecen: aparece como curso de doctorado en algunas Universidades –y siempre dentro del programa de Lingüística, dentro del área de conocimiento de Lingüística General; también se oferta en algunos cursos optativos o genéricos (fuera de cualquier plan curricular) como el que imparto desde hace varios años en la Universidad Complutense de Madrid: *Informática y textos literarios*<sup>1</sup>; y sólo una Universidad en España, la de Deusto en el País Vasco, a partir del 2004 concede un título oficial sobre esta materia: *Lenguas Modernas y Nuevas Tecnologías de la Información*, gracias al esfuerzo de Carmen Isasi y de Joseba Abaitua.

Así, la difusión de la Informática Humanística en el ámbito hispánico está limitada por su escasa presencia en los estudios académicos –que tiene su repercusión en la falta de interés por parte de las editoriales, tanto para publicar trabajos originales como para traducir textos importantes escritos en otras lenguas; de este modo, la difusión de la Informática Humanística se deja en manos del entusiasmo y el rigor personal –casi como un hobby– de algunos docentes. Y ya que estamos en Medievalia, no puedo dejar de recordar uno de estos ejemplos, que nos toca a

---

<sup>1</sup> <http://www.ucm.es/info/romantica/informatica.htm> (28 de abril de 2005)

todos muy cerca: la labor de Lilian von der Walde Moheno, que ha creado el portal *Hispanomedievalismo (Recursos en línea. Literatura)*<sup>2</sup>.

De este modo, la *Informática Humanística* (y la *Textual*, en concreto) se ofrecen como un campo virgen para el futuro, un campo nuevo de análisis y de experimentación donde los filólogos tenemos mucho que decir, sobre todo si no damos la espalda a las tecnologías informáticas a la hora de abordar nuevos métodos de trabajo y de difusión de los textos; cambios y retos revolucionarios como los que vamos asimilando y disfrutando en nuestra vida cotidiana, impensable hace tan solo unos años.

### 3. El texto no tienen quien le escriba

Las nuevas tecnologías informáticas presentan al texto un doble reto: de difusión y de análisis.

Se ha creado un nuevo medio de transmisión: las redes informáticas, que tienen en Internet y en las 3W su máximo exponente. Ante este nuevo medio los restos se agolpan esperando una respuesta:

¿Cómo difundir los textos –en nuestro caso, medievales- en la Red? ¿Habría que establecer modelos diferentes según la procedencia o la cronología de los mismos?

¿Qué estructuras han de mantenerse en este nuevo medio del medio de transmisión que los ha visto nacer y difundirse a lo largo de los siglos –el manuscrito o el impreso?

¿Cómo se concreta este modelo de hipertexto, es decir, de red de información, cuando queremos editar un texto medieval? ¿Hay diferencias de modelos hipertextuales si editamos un documento, un texto transmitido en una compilación, un impreso o un texto escrito y difundido en el siglo XIX o en el siglo XX?

¿Cuántos modelos y de qué tipo de edición podemos hablar en el ámbito informático?

¿Sería conveniente mantener los tipos de edición clásicos: transcripción paleográfica, presentación crítica o edición, ya que todos estos materiales pueden hacerse virtuales y ser sólo “reales” en el momento de nuestra lectura o utilización de un hipertexto?

Y en cuanto al reto del uso de las nuevas tecnologías informáticas para creación de herramientas que nos permitan un mejor análisis de los textos, las preguntas se multiplican. En este campo, podríamos distinguir, a su vez, dos ámbitos de estudio:

- a) Herramientas que ayuden en los distintos pasos para la realización de una edición crítica (*fontes criticae, recensio, examinatio...*). En este sentido, como ha demostrado el neolachmannismo, cuantos más datos pueda tener el editor, más posibilidades tiene de acertar en su *indicium*; y estos datos podrás resolver las mil dudas que conlleva toda edición.
- b) Herramientas que ayuden en el análisis de los textos ya editados, tanto en un plano textual, como literario o lingüístico.

Pero en los grandes proyectos del mundo hispánico –y, como no podía ser de otro modo, también en los pequeños, reflejos de los anteriores- que tienen hoy en día el *texto* como protagonista, como objeto de estudio dentro del ámbito de la informática, se han hecho de espaldas a la filología y a las múltiples posibilidades del hipertexto y de la informática humanística.

Veamos un ejemplo, paradigmático por ser la Biblioteca Virtual en español más visitada y más conocida: La *Biblioteca Virtual Miguel de Cervantes*<sup>3</sup> que fue creada en 1999. En estos últimos años ha mejorado su organización desde la Universidad de Alicante, así como su arquitectura y filosofía, aunque las huellas del pasado siguen presentes en los textos que hoy en día pueden aún consultarse. En un primer momento, el límite cronológico que se fijó para los

<sup>2</sup> <http://www.waldemoheno.net/recursos.html> (28 de abril 2005)

<sup>3</sup> <http://www.cervantesvirtual.com> (28 de abril 2005)

textos que debían incluirse se marcó según un único criterio: el necesario para no tener que pagar derechos de autor. Los textos medievales y de los Siglos de Oro fueron digitalizados a partir de ediciones en el mercado, sin indicación de su procedencia para así no tener que pagar derechos a los editores o a las editoriales, mostrando un gran desprecio por el trabajo científico que realizamos los filólogos. Dado el fracaso de este primer proyecto (debido a las protestas de la comunidad científica y de CEDRO, que no debieron gustar mucho a la Fundación Botín, que estaba detrás de su financiación), se optó por elegir ediciones “sin problemas” y trabajar con ejemplares que estuvieran en algún centro bibliográfico alicantino. Menos mal que esta fase parece ya concluida y se está potenciando la calidad gracias a las *Bibliotecas de Obras y Autores Clásicos*. Y así, en la biblioteca uno puede leer el *Libro de Buen Amor*, pero lo hará a partir de la edición impresa en París, por Louis-Michaud, sin año de publicación<sup>4</sup>. Por otro lado, la edición se ofrece de manera lineal, sin aprovechar ninguna de las posibilidades del hipertexto. Como resultado final se ofrece una de las primeras posibilidades de la relación del texto con las nuevas tecnologías: la digitalización de obras ya impresas en formato analógico. Se trabaja sólo siguiendo dos criterios: accesibilidad y acumulación. ¿Acaso que la información sea más accesible –siempre que se tenga ordenador y conexión a Internet- y se acumule una mayor cantidad es sinónimo de conocimiento? Si un mundo tuviera la accesibilidad a un ordenador y la paciencia de acumular horas tecleando quizás terminara por escribir el *Quijote*. Así lo pensaba Borges (con alguna que otra variante) y así de la mano del autor argentino nos topamos con la mejor caricatura de este tipo de proyectos que se basan exclusivamente en la cantidad: número de textos incorporados, número de visitantes diarios...

También es cierto que estos ejemplos son residuales en la Biblioteca Virtual Miguel de Cervantes, que en los últimos años está haciendo un gran esfuerzo dentro de la sección *Biblioteca de Obras y Autores clásicos*, en los que se trabaja para ofrecer en la red información textual de calidad, como muestran los portales dedicados a la *Celestina*, *Tirant lo Blanc* o *Martin Codax*, por no salirnos del ámbito medieval.

Las bibliotecas virtuales suponen el mejor ejemplo de las posibilidades actuales que las nuevas tecnologías ponen en el presente a nuestros intereses de difusión y edición de los textos. Aunque no en todas ellas se cumplen las mismas características, siendo escasas las que podemos definir como hipertextuales (Lucía Megías, 2005). Quedémonos con un ejemplo, donde se podrá apreciar las posibilidades actuales de la informática... y la falta de resultados científicos al hacerse la difusión de los textos al margen de la filología.

La biblioteca virtual *IntraText*<sup>5</sup> lleva colocando textos en la red desde 2001. En la actualidad (abril del 2005) pueden consultarse 5770 títulos, de 37 lenguas diferentes. Los textos proceden de otros proyectos de la Red, lo que propicia una de sus primeros elementos negativos: la heterogeneidad de la procedencia de las ediciones digitalizadas, en el caso que se indiquen, y la falta de información sobre los criterios ecdóticos que las sustentan, o sobre las ediciones analógicas que se encuentran en su base. Como *IntraText* sobresale más por las enormes posibilidades de lectura y análisis que pone a nuestra disponibilidad gracias a las herramientas informáticas que por la calidad textual del material incluido.

Cada texto puede ser leído y analizado desde diferentes perspectivas (ponemos ejemplos de la edición del *Quijote* de Miguel de Cervantes<sup>6</sup>):

a) Texto: secuencia lineal, mediante nodos de información vinculados entre sí (lo que es lo habitual –y único- en la mayoría de los proyectos textuales que hemos analizado en la Red). Las palabras, a excepción de las “formas vacías”, están vinculadas a su concordancia, lo que facilita el uso de la misma (Imagen 1).

<sup>4</sup> <http://www.cervantesvirtual.com/servlet/SirveObras/01479403120292341921157/index.htm> (28 de abril de 2005)

<sup>5</sup> <http://www.intratext.com/> (28 de abril de 2005)

<sup>6</sup> <http://www.intratext.com/y/ESL0027.HTM> (28 de abril de 2005)



Imagen 1: Texto: prólogo de la primera parte del *Quijote*

- b) Listas de palabras: se ofrecen varias posibilidades:
- b.1. Por orden alfabético (Imagen 2)
  - b.2. Por Frecuencia
  - b.3. Por orden inverso
  - b.4. Por longitud



Imagen 2: Lista de palabras por orden alfabético: *Don Quijote*

- c) Concordancia: no sólo se ofrece la concordancia, sino la posibilidad de acceder directamente a otras herramientas: listas alfabética y por frecuencia, concordancias de otras palabras, el texto, etc. (Imagen 3)



| Índice Ayuda    | Alfabetica [« »] | Frecuencia [« »]  | Don Quijote de la Mancha<br>IntraText - Concordancias |
|-----------------|------------------|---|---|
| Impresión       | lechones 1       | 9 lago  | <b>lector</b>   |
| Biblioteca      | lechos 2         | 9 lanzarote   |   |
| IntraText       | lección 1        | 9 leñimado  |   |
| Éulogos         | lector 9         | 9 lector  |   |
|                 | lectura 2        | 9 lequiste  |   |
|                 | ledañas 1        | 9 letrados  |   |
|                 | lee 8            | 9 levantada   |   |
| Parte. Capitulo |                  |   |   |
| 1               | I. Pr            | Prólogo~ Desocupado lector, sin juramento me podrás         |   |
| 2               | I. Pr            | ojos, como otros hacen, lector carísimo, que perdone o      |   |
| 3               | I. Pr            | prólogo, en el cual verás, lector suave, la discreción de   |   |
| 4               | I. 20            | y eran (si no lo has ¡oh lector!, por pesadumbre y enojo)   |   |
| 5               | I. 45            | espacio, porque no era buen lector, a cada palabra que leía |   |
| 6               | II. Pr           | Prólogo al lector~ ¡Válame Dios, y con                      |   |
| 7               | II. Pr           | de estar esperando ahora, lector ilustre, o quier plebeyo,  |   |
| 8               | II. Pr           | no le cuadrare, dirásle, lector amigo, éste, que, también   |   |
| 9               | II. 44           | con pucheritos.~ Deja, lector amable, ir en paz y en hora   |   |

Imagen 3: Concordancia: palabra “lector”: *Don Quijote*

d) Estadísticas: se ofrece como una visión global de los datos analizados en las diferentes secciones anteriores. Para el caso del *Quijote*, se ofrecen los siguientes resultados:

- Concurrencias: 377767
- Palabras (tokens): 22654
- Conc./Palabras: 16,68
- Concurrencia de palabras de función: 198347
- Palabras de función: 207
- Longitud media de concurrencia: 4,30
- Longitud media de las palabras: 7,84

El proyecto *Intratext* se ha realizado a partir de un programa de concordancias creado por Éulogos, al que denominan ECP (*Éulogos Concordance Program*), y las indicaciones de la estructura (lo que permite los enlaces y las diferentes listas) se han realizado con marcadores ETLM (*Éulogos Text Markup Language*). Y, sobre todo, se ha pretendido su uso universal, con lo que tanto los nodos como los enlaces se han realizado para que sean compatibles con todos los sistemas operativos y para que puedan correr en cualquier programa de navegación.

El gran despliegue técnico de *IntraText*, así como las útiles herramientas informáticas que pone a la disposición del investigador, permite volver –una y otra vez- al aspecto más criticable de todo el proyecto y de tantos otros que hemos analizado en estas páginas: la escasa validez científica del material textual. Si los proyectos que tienen al texto como centro y objeto de estudio han de establecer sus bases en tres ámbitos de estudio: tecnología informática, lingüística computacional y filología, *IntraText*, como suele suceder con tantos otros proyectos de la Red, ha prestado especial atención a los primeros –como hemos tenido ocasión de reseñar en las páginas anteriores- pero ha desestimado el último: ¿qué *texto* del *Quijote* se ha utilizado para la realización de las listas, de la concordancia, de las estadísticas....? ¿Qué interpretación filológica es la que está en la base? ¿La transcripción de un testimonio impreso determinado, de cuál en concreto? ¿O tal vez la “hipótesis de trabajo” de un determinado editor, cuyo método de trabajo, cuyas hipótesis concretas se omiten al silenciar su nombre? Al entrar en el apartado de los “créditos”, la información no puede ser más escueta (y más decepcionante):

|  |                      |
|--|----------------------|
| Fuente impresa                         | Edición I: 1605-1615 |
| Fuente de la transcripción electrónica | Internet             |

Marcador ETML

Redacción IntraText

Detalles de publicación

¿Es esta la información textual que esperaríamos de un proyecto que ha cuidado con tanto mimo las herramientas informáticas y las posibilidades hipertextuales que ofrece este nuevo medio de transmisión, al vincular la información de una manera tan rica? En absoluto. Una vez más, el déficit de ciencia en el ámbito textual, resta credibilidad a un proyecto que, desde otras perspectivas, resulta magnífico.

Las herramientas informáticas sigue siendo el campo de mejora en el prototipo experimental de una biblioteca *IntraText Plus*<sup>7</sup>, que en la actualidad ofrece el texto de la *Divina Comedia*. Las ventajas frente al modelo hipertextual ofrecido por *IntraText* se han centrado en la capacidad de acceder, de una manera más eficiente y rápida a todas las posibilidades de análisis y lectura que un texto en formato hipertextual posibilita, como se aprecia en la siguiente imagen (Imagen 4):

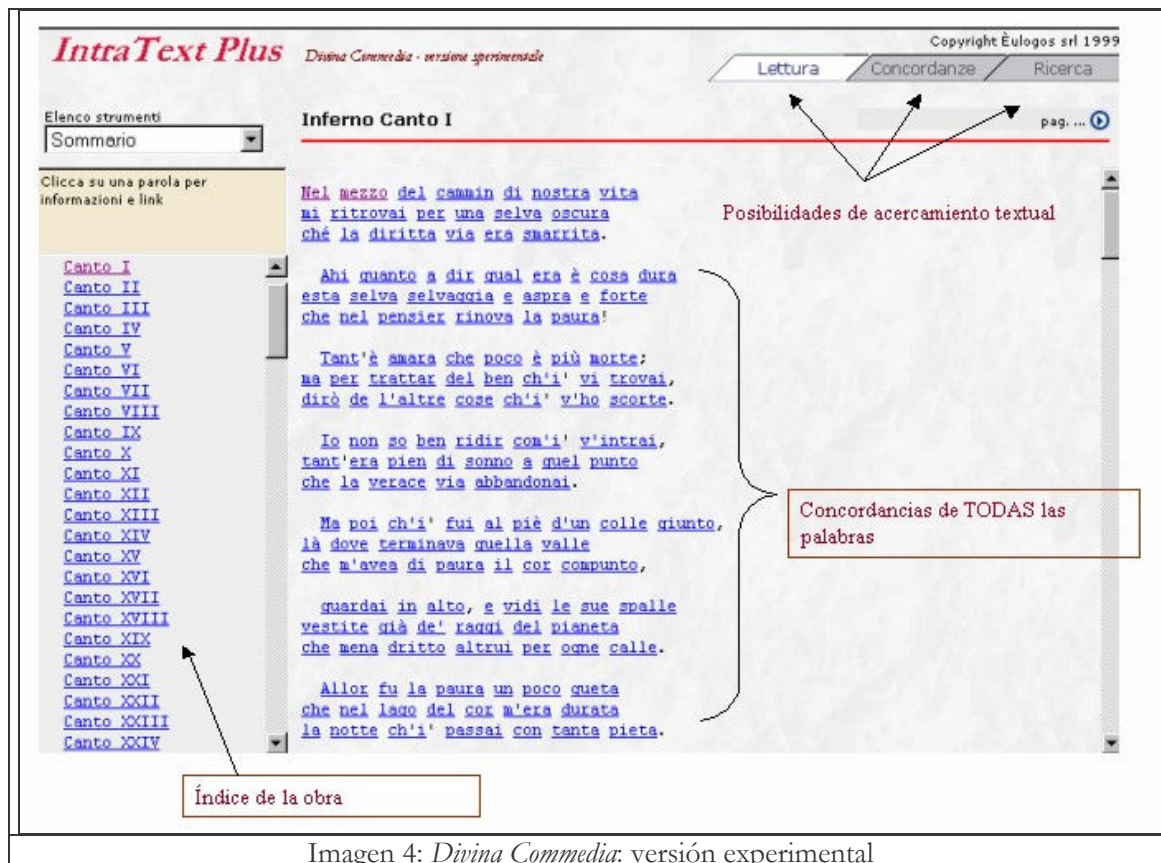


Imagen 4: *Divina Comedia*: versión experimental

a) TEXTO: en el lateral derecho se ha insertado una serie de enlaces estructurales que permiten moverse por el texto sin tener que cambiar de interfaz. Por otro lado, en la esquina superior derecha se permite seguir avanzando mediante su secuencia lineal.

<sup>7</sup> <http://www.intratext.com/IXT+/>

b) CONCORDANCIAS: se realiza de todas las palabras; además, se incorpora en la misma interfaz la lista alfabética de palabras, así como de frecuencia, un buscador y la posibilidad (todavía no activada) de realizar la concordancia por lemas y no por formas (imagen 5).

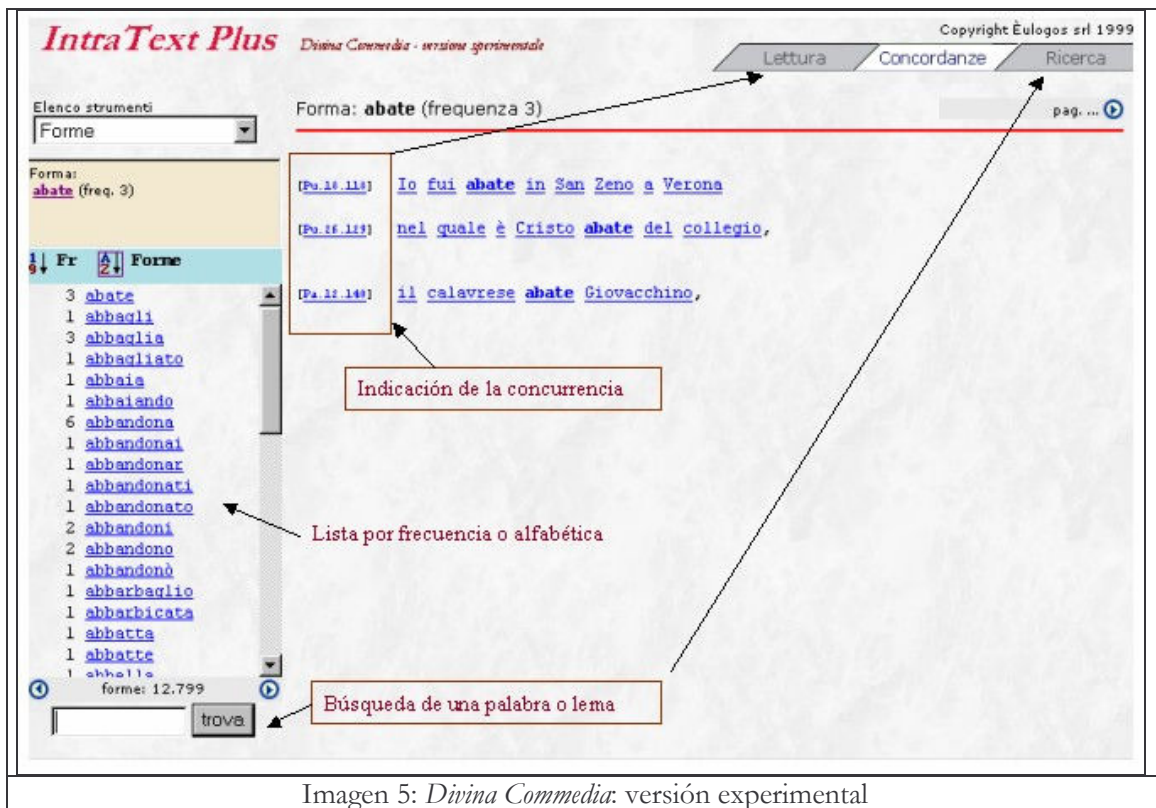


Imagen 5: *Divina Commedia*. versión experimental

c) MOTOR DE BÚSQUEDA: frente a las listas y las estadísticas (que aparecen dentro de la opción de *Concordancias*, se ha incorporado una nueva herramienta: un motor de búsqueda. La petición de información puede limitarse mediante tres parámetros: [a] Cántica, [2] Canto, [3] verso.

En el explorador izquierdo, se anuncian otros modos de análisis y de estudio, que ahora no están activados, como la opción de realizar lecturas interlineales de un texto en varios idiomas diferentes.

Las posibilidades que ofrecen las nuevas tecnologías para el estudio de los textos literarios abren nuevas vías de investigación y de análisis hasta ahora impensables, hasta ahora imposibles dentro de las limitaciones espaciales y memorísticas de la mente humana. Pero para que estas herramientas ofrezcan sus verdaderos frutos, para que el resultado de su uso pueda ser considerado realmente científico (y por tanto, fiable para la realización de hipótesis de interpretación y de estudio), es necesario que el material de base, el material textual haya pasado el control de la filología, la única ciencia que es capaz de plantear las hipótesis de trabajo científicas sobre los múltiples problemas que plantea. Y estos problemas han de ser resueltos (o al menos, indicados) antes de poner a la ingeniería informática a trabajar. La *Informática Textual* tiene que basarse en textos fiables y no, exclusivamente, en excelentes programas informáticos, en ambiciosas interfaces de diseño sencillo y efectivo. Una vez más se pone de manifiesto cómo los proyectos que reúnen textos y nuevas tecnologías no pueden prescindir de la *Informática Textual*, de la filología, la única ciencia que tiene como objeto de estudio el texto.



#### 4. ¿Crónica de un fracaso anunciado?

Aquí están delante de nuestros ojos las nuevas posibilidades: no se trata tanto de futuro, de soñar con nuestras herramientas, como de ser capaces de aceptar el pulso del presente.

Herramientas que ofrecen una acumulación de la información hasta ahora impensable; herramientas que permiten una recuperación de la misma en décimas de segundo. Los límites parecen haber explotado. ¿Pero con qué finalidad?

Frente a lo que sucede con el otro gran corpus académico, el CREA, *Corpus de Referencia del Español Actual*, que va desde 1975 hasta nuestros días, se utilizan tantos textos escritos como orales, de una gran variedad de procedencias<sup>8</sup>, el CORDE<sup>9</sup> está compuesto esencialmente de textos escritos, de textos transmitidos por canales tan diferentes como el manuscrito, la imprenta manual, la moderna o los pre-textos. En la actualidad, se acumulan más de 250 millones de registros, y se anuncia que para finales de 2005 se corregirán los 300 millones de registros de los que actualmente consta el proyecto. Pero esta herramienta informática, la mejor de todas las realizadas para el ámbito hispánico, que pretende reunir todos los textos hispánicos desde los orígenes hasta 1974, ¿por qué no se ha convertido en un referente de la filología hispánica? Las razones pueden ser múltiples, pero entre todas ellas destaca el hecho de que se ha hecho de espaldas a la filología, aunque se haya beneficiado la Real Academia Española de millonarios presupuestos gracias a los programas I+D (Investigación y Desarrollo).

El CORDE cuenta con un buen motor de búsqueda, que permite recuperar la información desde diferentes posibilidades: Documentos, Concordancias, Párrafos y Agrupaciones, que permiten obtener en un momento un enorme caudal de información para llevar a cabo estudios lingüísticos. Así como permite acceder a otras dos herramientas que pueden ayudarnos en estudios lingüísticos, como son las Estadísticas (permite conocer cómo un registro se distribuye en el arco diacrónico de su documentación) y los Filtros (permiten afinar mucho más las búsquedas según nuestras necesidades científicas).

Desde la *Informática Textual* son tres los aspectos discutibles que hacen que los resultados del CORDE se tengan que contrastar con otras fuentes, lo que dificulta su uso (en algunos casos) y lo hace poco fiable en otros muchos. Algunos de estos aspectos se explican desde la propia historia de la filología hispánica, que no supo en el siglo XX desarrollar una verdadera ciencia editorial, por lo que hay una carencia de “hipótesis de trabajo científicas” (de ediciones críticas, en otras palabras) que permitan el conocimiento de gran número de textos medievales y renacentistas; pero otros podrían, de manera no muy complicada, corregirse en el actual diseño del proyecto.

1. La AUTORIDAD. En un momento anterior, se ha hecho alusión a las tres modalidades de las que, según el propio proyecto, proceden los textos: se hablaba de libros escaneados y de otros que se habían conseguido ya en formato electrónico, sin olvidar unos pocos, que por peculiaridades del lenguaje –y también por la mala calidad del papel, como la prensa del siglo XIX– se habían teclado. Pero, ¿qué ediciones se utilizan? ¿Qué criterio científico se ha utilizado a la hora de escanear, a la hora de aceptar textos digitalizados –como los del proyecto del Hispanic Seminary of Medieval Studies de la Universidad de Madison–, o de teclear algunos de ellos? ¿El criterio es que se trate de una edición accesible, que la haya realizado un académico, que se encuentre en la biblioteca del centro? ¿Qué criterio se ha seguido para descartar una determinada edición, una determinada hipótesis de trabajo de un particular científico? En ningún momento se hacen públicos estos datos, lo que es necesario; pero lo peor es que da la impresión que tampoco en ningún momento se han clarificado en el diseño del proyecto a los criterios que han llevado a

---

<sup>8</sup> Tal y como se indica en el portal de Internet de la Real Academia Española: “Los textos escritos, procedentes tanto de libros como de periódicos y revistas, abarcan más de cien materias distintas. La lengua hablada está representada por transcripciones de documentos sonoros, procedentes, en su mayor parte, de la radio y la televisión”.

<sup>9</sup> <http://www.rae.es> (28 de abril de 2005)

los responsables del corpus a elegir una determinada edición por encima de otra. Analicemos un ejemplo concreto para comprender, en todas sus consecuencias, la falta de autoridad de un proyecto cuando se ha primado la cantidad (¡180 millones de registros en la actualidad!) frente a la calidad de los textos —y por *texto* entendemos la última voluntad del autor, a la que sólo se puede acercar uno como hipótesis de trabajo científica, como ha demostrado la crítica textual.

Volvamos a la búsqueda de una determinada forma (*fazjenda*), tal y como ya hice en *Filología Románica en Internet. I. Los textos* (2002) cuando se comentaba este mismo problema. Al buscar *fazjenda*, limitando la búsqueda a una época concreta (1200-1500) y al tema 12. *Narrativa*, se obtienen el siguiente resultado: 403 casos en 28 documentos, que son los siguientes, según los datos aportados por el mismo proyecto:

|    | Año         | Autor                       | Obra   | Tema  | Publicación  |
|----|-------------|-----------------------------|--|---|--|
| 1  | a 1482      | Anónimo                     | <i>Esopete ystoriado. Toulouse, Jobann Paris, 1488. Manchester, John Rylands Library</i> | 12.Relato breve culto                             | Victoria A. Burrus, Hispanic Seminary of Medieval Studies (Madison), 1995                |
| 2  | 1427 - 1428 | Villena, Enrique de         | <i>Traducción y glosas de la Eneida. Libros I-III</i>                                    | 12.Relato extenso novela y otras formas similares | Pedro M. Cátedra, Turner Libros (Madrid), 1994   |
| 3  | 1293        | Anónimo                     | <i>Gran Conquista de Ultramar. Ms. 1187 BNM</i>  | 12.Relato extenso novela y otras formas similares | Louis Cooper; Franklin M. Waltman, Hispanic Seminary of Medieval Studies (Madison), 1995 |
| 4  | c 1450      | Anónimo                     | <i>Traducción de la Teseida de Boccaccio</i>   | 12.Relato extenso novela y otras formas similares | Michael Gerli, Hispanic Seminary of Medieval Studies (Madison), 1991                     |
| 5  | 1438        | Martínez de Toledo, Alfonso | <i>Arcipreste de Talavera (Corbacho)</i>   | 12.Otras formas                                   | Marcella Ciceri, Espasa-Calpe (Madrid), 1990   |
| 6  | c 1253      | Anónimo                     | <i>Sendebar</i>  | 12.Relato breve tradicional                       | María Jesús Lacarra, Cátedra (Madrid), 1989  |
| 7  | c 1500      | Anónimo                     | <i>Historia de la reina Sebilla</i>  | 12.Relato extenso novela y otras formas similares | Nieves Baranda, Turner Libros (Madrid), 1995   |
| 8  | c 1470      | Anónimo                     | <i>La demanda del Sancto Grial</i>   | 12.Relato extenso novela y otras formas similares | Adolfo Bonilla y San Martín, Bailly-Bailliére (Madrid), 1907                             |
| 9  | c 1250      | Anónimo                     | <i>Poridat de poridades. Escorial L.III.2</i>  | 12.Relato breve tradicional                       | Pedro Sánchez-Prieto Borja, Universidad de Alcalá de Henares (Alcalá de Henares), 2004   |
| 10 | c 1325      | Anónimo                     | <i>Historia del caballero Plácidas. Escorial b.I.13</i>                                  | 12.Relato extenso novela y otras formas similares | Pedro Sánchez-Prieto Borja, Universidad de Alcalá de Henares (Alcalá de Henares), 2004   |
| 11 | c 1500      | Anónimo                     | <i>Libro del conde Partinuplés</i>   | 12.Relato extenso novela y otras                  | Nieves Baranda, Turner Libros (Madrid), 1995   |

X JORNADAS MEDIEVALES (Del 20 al 24 de septiembre del 2004)

|    |                 |                              |   |   |  |
|----|-----------------|------------------------------|---|---|--|
|    |                 |                              |   | formas similares                                  |  |
| 12 | c 1300 - 1325   | Anónimo                      | <i>Cuento muy fermoso de Otas de Roma</i>   | 12.Relato extenso novela y otras formas similares | Herbert L. Baird, Jr., Real Academia Española (Madrid), 1976                               |
| 13 | 1499            | Anónimo                      | <i>La historia de los nobles caballeros Oliveros de Castilla y Artús d'Algarbe</i>                          | 12.Relato extenso novela y otras formas similares | Nieves Baranda, Turner Libros (Madrid), 1995   |
| 14 | c 1400 - 1498   | Anónimo                      | <i>El baladro del sabio Merlin con sus profecías</i>  | 12.Relato extenso novela y otras formas similares | Isabel Hernández González, CILUS (Salamanca), 1999   |
| 15 | 1325 - 1335     | Manuel, Juan                 | <i>El Conde Lucanor</i>   | 12.Relato breve culto                             | Guillermo Serés, Crítica (Barcelona), 1994   |
| 16 | c 1414          | Anónimo                      | <i>Traducción de Lanzarote del Lago</i>   | 12.Relato extenso novela y otras formas similares | Harvey Sharrer, Hispanic Seminary of Medieval Studies (Madison), 1999                      |
| 17 | c 1500          | Anónimo                      | <i>La crónica del noble cavallero el conde Fernán Gonçales, con la muerte de los siete infantes de Lara</i> | 12.Relato extenso novela y otras formas similares | Nieves Baranda, Turner Libros (Madrid), 1995   |
| 18 | c 1450          | Anónimo                      | <i>Ejemplos muy notables</i>  | 12.Relato breve culto                             | Silvia Iriso Ariz, Lemir (Valencia), 2001  |
| 19 | 1313 - 1469     | Anónimo                      | <i>La estoria de Merlin</i>   | 12.Relato extenso novela y otras formas similares | Karl Pietsch, The University of Chicago (Chicago), 1924                                    |
| 20 | 1498            | Anónimo                      | <i>Historia de Enrique fijo de doña Oliva, rey de Jerusalem y emperador de Constantinopla</i>               | 12.Relato extenso novela y otras formas similares | Nieves Baranda, Turner Libros (Madrid), 1995   |
| 21 | 1458            | Díaz de Toledo, Pero         | <i>Diálogo é razonamiento en la muerte del marqués de Santillana</i>  | 12.Relato extenso diálogo y miscelánea            | Antonio Paz y Melia, Sociedad de Bibliófilos Españoles (Madrid), 1892                      |
| 22 | 1300 - 1305     | Anónimo                      | <i>Libro del cavallero Cifar</i>  | 12.Relato extenso novela y otras formas similares | Juan Manuel Cacho Bleuca, Universidad de Zaragoza (Zaragoza), 2003                         |
| 23 | c 1400 - c 1421 | Sánchez de Vercial, Clemente | <i>Libro de los exemplos por A. B. C.</i>   | 12.Relato breve culto                             | John Esten Keller, CSIC (Madrid), 1961   |
| 24 | 1251            | Anónimo                      | <i>Calila e Dimna</i>   | 12.Breve  | Juan Manuel Cacho Bleuca; María Jesús Lacarra, Castalia (Madrid), 1993                     |
| 25 | 1425 - 1450     | Rodríguez del Padrón, Juan   | <i>Bursario</i>   | 12.Relato extenso novela y otras formas similares | Pilar Saquero Suárez-Somonte; Tomás González Rolán, Universidad Complutense (Madrid), 1984 |
| 26 | a 1250          | Anónimo                      | <i>Bocados de oro</i>   | 12.Relato breve tradicional                       | Mechthild Crombach, Romanisches Seminar der  |

|    |             |                              |  |  |  |
|----|-------------|------------------------------|--|--|--|
|    |             |                              |  |  | Universität Bonn (Bonn), 1971                    |
| 27 | c 1480      | San Pedro, Diego de          | <i>Tractado de amores de Arnalte y Lucenda</i> | 12. Relato extenso novela y otras formas similares | Keith Whinnom, Castalia (Madrid), 1985           |
| 28 | 1482 - 1492 | Rodríguez de Montalvo, Garcí | <i>Amadís de Gaula, libros I y II</i>          | 12. Relato extenso novela y otras formas similares | Juan Manuel Cacho Blecua, Cátedra (Madrid), 1991 |

En este listado de quince ediciones, se documenta una alarmante heterogeneidad de ediciones, tanto por sus características como por sus finalidades, que hace imposible –o al menos así nos lo enseña la filología– la comparación de sus resultados, lo que sí se espera al incluirlas todas en un mismo banco de datos, sin ningún criterio diferenciador entre ellos –lo que sí que permitiría la tecnología informática, por otro lado. De este modo, encontramos:

- a) Las transcripciones de un determinado testimonio, como así sucede con los procedentes del proyecto del Hispanic Seminary of Medieval Studies de Madison, en donde se seguía un particular sistema de transcripción, pensado para su posterior uso informático con la finalidad de realizar un diccionario del español medieval, y nunca para su lectura como libro, a pesar de que hayan sido muchos los textos que sólo están disponibles en un acercamiento editorial moderno gracias a esta colección. Dentro de este grupo se engloba el nº 3: la transcripción del ms. 1187 de la Biblioteca Nacional de Madrid, que ha conservado sólo los últimos ciento treinta y cinco capítulos del tercer libro y el cuarto completo de la *Gran Conquista de Ultramar* (DFLME, 2002: 603-608); o el *Esopete ystoriado*, editado por Victoria A. Burrus (nº 1).
- b) Ediciones modernas, que se han publicado en colecciones donde se prima la transcripción crítica de un testimonio manuscrito, como es la Biblioteca Castro: nº 2, 7, 11 y 2'; o en nº 8, que es también una transcripción del incunable de uno de los primeros textos caballerescos.
- c) Ediciones modernas, publicadas en editoriales y colecciones de prestigio (como Castalia, Cátedra, Crítica o Espasa-Calpe), que ofrecen una gran variedad de posibilidades de acercamiento textual, desde los más apegados a un determinado testimonio, a los que se acercan al estudio completo de su transmisión: nºs 5, 15, 27 y 28.
- d) Ediciones críticas, en donde los investigadores ofrecen una determinada hipótesis de trabajo, después de analizar la transmisión completa de una determinada obra: nº 25.
- e) Ediciones realizadas a finales del siglo XIX o principios del siglo XX (nºs 8 y 21), con unos criterios científicos muy alejados –en ocasiones– de los mínimos presupuestos ecdóticos actuales.

Lo cierto es que la ausencia de una tradición ecdótica, de la escuela que sea, en España, hace dificulta, en gran medida, el diseño y el desarrollo proyectos textuales como el presente, de una ambición desbordante, lo que no sucede en otros países como Italia o Francia, al margen de que en un caso nos movamos en la línea certera del lachmannismo y del neo-lachmannismo, y en el otro en el arte propugnado por Bédier, que ha terminado por primar la transcripción crítica de testimonios. Pero la dificultad de partida –enorme, repito– no justifica la mezcla de “autoridad” a la hora de trabajar con ediciones: no es suficiente con haber escaneado bien un libro, con haberlo tecleado adecuadamente, con haberlo marcado y etiquetado de manera correcta, con haber conseguido que los programas informáticos lo reconozcan –labor que compete, sobre todo, a los técnicos informáticos–; además es necesario, es necesario sobre todo que estos libros hayan sido seleccionados, o hayan sido transcritos o editados, de acuerdo a unos criterios filológicos, que son



los que marcarán su “autoridad”, como la *Informática Textual* ha venido enseñando –y lo seguirá haciendo- en los últimos años.

Del listado anterior hemos dejado fuera varias entradas, que merecen un comentario final, por ser uno de los aspectos novedosos que se van introduciendo en las últimas versiones del CORDE<sup>10</sup>: el nº 11, que se corresponde con el texto del *Libro del caballero Zifar*, con la siguiente indicación: “Juan Manuel Cacho Bleuca, Universidad de Zaragoza (Zaragoza), 2003”, y lo mismo puede decirse de los nº 9 y 10: “Pedro Sánchez-Prieto Borja, Universidad de Alcalá de Henares (Alcalá de Henares), 2004”. En la presentación que en Internet se hacía de la mejora que se terminó en octubre de 2003, de donde arrancan estas mejoras textuales, se destaca una mayor preocupación por la calidad textual de los materiales que se incorporan al proyecto: “Además del considerable incremento del volumen de textos, la nueva versión ha sustituido algunas ediciones por otras más adaptadas a los niveles de calidad habituales en las ediciones actuales”. En la anterior versión, como ya se indicó en otra publicación (Lucía Megías, 2002: 75), se utilizó para introducir en el CORDE el texto del *Zifar* (principios del siglo XIV), la transcripción que Francisco Gago Jover había realizado de un testimonio manuscrito de finales del siglo XV (ms. P de la Bibliothèque Nationale de France), para el HSMS de Madison. La sustitución por una edición de Juan Manuel Cacho Bleuca no puede ser más acertada: pocos investigadores pueden encontrarse hoy que conozcan mejor el texto del *Zifar*. Pero, ¿qué criterios ha utilizado Juan Manuel Cacho Bleuca para su edición? ¿La transcripción de uno de los tres testimonios conservados? ¿La edición crítica del texto? En el caso de las ediciones publicadas, siempre existe la posibilidad de consultarlas en la biblioteca, pero ¿qué sucede con estos “textos”, que nacen del acuerdo de colaboración entre la Real Academia Española y diferentes universidades y centros científicos de España? No me cabe duda, de que se ha ganado en calidad textual, pero no así en el aspecto de “autoridad” que aquí se viene comentando.

En la nueva versión del Corde, de marzo de 2005, se ha incorporado a los bancos de datos de la RAE una herramienta que habíamos indicado en anteriores trabajos: el listado completo de los textos que, en cada momento, están incorporados al CORDE, con la referencia bibliográfica pertinente, ya sea la edición de la que se ha escaneado, ya sea el proyecto científico que ha presentado el texto ya digitalizado. Y esta “Nómina de autores y obras” se ha convertido en un motor de búsqueda, con un interfaz de fácil manejo, como suele ser habitual en los proyectos de la RAE (imagen 6)

REAL ACADEMIA ESPAÑOLA  
Jueves, 28 de abril de 2005

Real Academia Española - Banco de datos del Español - Nómina de autores y obras

Criterios de selección:

Corpus: TODOS ● CREA ● CORDE ●

Autor:

Obras:

Cronológico:

Medio: (Todos) Libros Prensa Miscelánea (solo CREA)

Geográfico: (Todos) Argentina Bolivia Chile Colombia Costa Rica

Buscar Limpiar

[Estadísticas generales](#)

Imagen 6: Nómina de autores y obras

Pero, teniendo en cuenta la carencia filológica de la ciencia hispánica durante el siglo XX, de la que ya nos hemos lamentado en estas páginas, se haría necesario acompañar esta “Nómina

<sup>10</sup> Aunque el CORDE es un proyecto que se va completando en el tiempo, sería muy útil –a la hora de poder ir precisando los cambios y mejoras que se van introduciendo- que se fuera indicando de algún modo las versiones por las que ha pasado, uso habitual en programas y en proyectos de esta envergadura, en donde la actualización supone una de sus grandes ventajas.

de autores y obras” de una notas filológicas, que expusieran, de una manera clara y concisa, [1] las características de la edición utilizada como base, [2] la problemática textual que rodea al texto – siempre que se considere oportuno-, y [3] clasificación de la edición utilizada, teniendo en cuenta una tipología previamente establecida, que va desde la edición crítica a la mera transcripción paleográfica de un determinado testimonio.

De esta manera, podríamos –desde el propio proyecto- ofrecer al usuario los datos necesarios para que pueda contrastar los resultados científicos obtenidos. No se trata –¡por supuesto!- de realizar ediciones nuevas de todos los textos sino de todo lo contrario: de utilizar lo que se ha hecho, de mejorarlo y de situarlo en su verdadero contexto científico y filológico. Una vez más, se muestra la necesidad de tener en cuenta la *Informática Textual*; una vez más, se muestra la necesidad de tener en cuenta las peculiares características de los *textos* a la hora de dar forma a las nuevas herramientas informáticas.

2. LA PRESENTACIÓN GRÁFICA: El diverso origen de las diferentes ediciones y transcripciones de las que se vale el proyecto para llegar a esas escalofrantes cifras de millones de registros, tiene también una consecuencia que limita la capacidad de búsqueda: la heterogeneidad de las presentaciones gráficas. La búsqueda de *fazienda* (entre 1200-1500 en obras narrativas) nos daba el resultado de 493 casos en 28 documentos... pero no es la única forma de esta palabra en el corpus: *fazienda* aparecen documentada una vez en la misma transcripción que Louis Cooper y Franklin M. Waltman realizaron para el Hispanic Seminary of Medieval Studies (Madison) en 1995, del ms. 1187 de la Biblioteca Nacional de Madrid, que ha transmitido *La Gran Conquista de Ultramar*, que documenta 61 casos de *fazienda*, así como 23 veces en la traducción del *Lanzarote del Lago*, según el ms. conservado en la Biblioteca Nacional de Madrid, que en 8 ocasiones documenta la forma *fazienda*.

Y así lo podríamos multiplicar con otros ejemplos, como sucede con el infinitivo *fazer*, que da resultados diferentes según la presentación gráfica, entre 1200-1500, en textos narrativos:

- a) *fazer*: 4563 casos en 43 documentos.
- b) *facer*: 29 casos en 5 documentos.
- c) *ffazer*: 6 casos en 3 documentos.
- d) *Fazer*: 293 casos en 3 documentos (265 casos en la traducción del *Lanzarote del Lago*, y los 28 restantes en el *Libro de los gatos*, según la transcripción paleográfica que llevó a cabo Bernard Darbord en 1984).

Este aspecto, que limita las enormes posibilidades que ofrece la tecnología informática para poder trabajar con cantidades tan abundantes de información, podría corregirse mediante dos procedimientos: adaptar a unas mismas normas de presentación gráfica todo el corpus (organizado, quizás, por épocas), como así lo realizó el proyecto del HSMS, que sí que ha sido volcado con sus modelos en el CORDE; o, lo que es preferible, en vez de trabajar con formas hacerlo con lemas, en donde se recogen todas las formas que lo han documentado, en sus particulares modos de presentación, lo que permitiría estudios grafemáticos y lingüísticos, que ahora no pueden realizarse con el proyecto en su formulación actual. La lematización y la etiquetación de los *corpora textuales* se ha convertido en uno de los requisitos imprescindibles de este tipo de proyectos. Al mismo tiempo, se podría incluir, además de las ediciones, transcripciones paleográficas de los testimonios, lo que permitiría avanzar en el estudio de la lingüística diacrónica.

3. LA CRÍTICA TEXTUAL. Por último, hay un error de base –lección de la crítica textual- que limita, en gran medida, el uso que podemos hacer de algunas de las herramientas informáticas que ofrece el CORDE: la diferencia entre *texto* y *testimonio*, entre *génesis* y *transmisión*. Este error hace inútiles, en gran medida, los datos que aparecen en herramientas como las estadísticas, que se

basan en el marcado cronológico que se ha hecho de cada texto... y este marcado cronológico (que aparece en una de las primeras columnas del listado anterior) se ha realizado sobre la génesis del texto, y no sobre la datación de los testimonios que se toman como texto base de las diferentes transcripciones o ediciones que se han utilizado en el proyecto. Veamos unos pocos ejemplos, siguiente el listado anterior:

a) nº 14: *El Baladro del sabio Merlín*: que se data con la fecha doble: c. 1400-c.1498: la primera no sabría muy decir a qué corresponde, ya que en la versión anterior del Corde se indicaba la fecha de 1313, que es la que crítica asigna a la génesis de la traducción, mientras que el único testimonio que lo ha conservado –manuscrito de la Biblioteca Universitaria de Salamanca– está datado en 1469; la segunda fecha, en cambio, no sería tanto “circa”, como el año 1498, cuando se termina de imprimir el incunable, que sirve de base a la transcripción de Isabel Hernández. ¿Por qué no marcar de alguna manera esta información? ¿Por que no diferenciar claramente lo que es la documentación del testimonio (1498) de lo que es el texto original, que se ha perdido, y cuya forma lingüística nada tendrá que ver con el incunable?

b) nº 19: *La Estoria de Merlín*: que sucede otro tanto, aunque en este caso sí que se mantiene “1313-1469”: la fecha de 1313 vuelve a hacer alusión a la génesis, pero el testimonio manuscrito –una verdadera compilación sapiencial, por otro lado– está fechado en 1469, que es a la fecha que deberíamos adscribir, en última instancia, la forma lingüística del texto tal y como aparece en el testimonio conservado.

c) nº 24: *Calila e Dimna*, que se data en 1251; la edición de Juan Manuel Cacho Bleuca y de M<sup>a</sup> Jesús Lacarra utiliza como texto base el ms. A, el que se conserva en la Biblioteca del Monasterio de El Escorial: ms. h-III-9, datado a principios del siglo XV. Estas son las palabras de los editores a la hora de justificar la elección de A como texto base: “De los medios externos se desprende que el manuscrito A es el más antiguo, y de un cotejo interno se puede deducir que el manuscrito A cuantitativamente está menos modernizado. Los resultados no son concluyentes por cuanto hay ciertas formas que en el manuscrito A, ocasionalmente, resultan más modernas. Por poner solo un caso, hay formas en –ié del imperfecto que se conservan en B y han desaparecido en A. Resulta sorprendente que un copista medieval hubiera mantenido del texto sin modernizarlo, cuando para él suele ser algo vivo y susceptible de adaptarse a los nuevos tiempos” (1993: 65). ¿Podemos seguir atribuyendo determinado léxico, determinadas formas del *Calila e Dimna* del manuscrito del siglo XV al 1251, fecha de su posible génesis?

d) nº 28: *Amadís de Gaula*: que se data según el texto refundido de Garci Rodríguez de Montalvo, o de su perdida primera edición de 1482-1492, cuando lo que hemos conservado –y lo que se edita– es la edición de 1508 que viera a la luz en lo talleres zaragozanos de Jorge Coci. ¿Debemos incluir *Amadís de Gaula* dentro de nuestra pesquisa entre 1200-1500, cuando las primeras documentaciones impresas se datan en 1508?

Veamos ahora los resultados que aparecen en la herramienta de “estadísticas” según la búsqueda de *fazienda* con los criterios anteriormente indicados:

| Estadísticas |       |       |        |        |       |                   |        |       |
|--------------|-------|-------|--------|--------|-------|-------------------|--------|-------|
| Año          | %     | Casos | País   | %      | Casos | Tema              | %      | Casos |
| 1251         | 32.82 | 86    | ESPAÑA | 100.00 | 403   | 11.- Prosa lírica | 100.00 | 403   |
| 1293         | 23.28 | 61    |        |        |       |                   |        |       |

|      |       |    |
|------|-------|----|
| 1482 | 12.97 | 34 |
| 1498 | 2.67  | 7  |
| 1300 | 1.90  | 5  |
| 1427 | 1.90  | 5  |
| 1499 | 1.90  | 5  |
| 1313 | 1.52  | 4  |

¿Debemos seguir adscribiendo, como ya se ha indicado, a 1251 los resultados del ms. del siglo XV del *Calila e Dimna*? ¿Por qué se toma la fecha de 1313 para los textos sobre la *historia de Merlín*, cuando pertenecen a un incunable de 1498 y a una copia de finales de la misma centuria? ¿Acaso las documentaciones (35) de 1482, pertenecientes al *Amadís de Gaula*, no deberíamos situarlas con todo rigor en 1508? Y las 9 documentaciones del *Libro del caballero Zifar*, que se datan en 1300, ¿pertenecen a uno de los dos testimonios manuscritos del siglo XV o al impreso de 1512?

La distinción entre *texto* y *testimonio*, entre génesis y transmisión, la marcación adecuada de cada uno de las fuentes de información del proyecto desde la perspectiva de la *Informática Textual*, podría solucionar este problema que limita, en gran medida, la utilización científica de las herramientas informáticas que pone a disposición de la comunidad científica el CORDE. Las etiquetas de los diferentes “textos” incluidos estaría estrechamente relacionada con las notas filológicas que deberían acompañar a los “documentos”, tal y como se ha indicado con anterioridad. De este modo, se podría –de acuerdo con el análisis de las ediciones utilizadas– realizar un doble etiquetado: el del texto (1251 en el caso del *Calila e Dimna*) y el del testimonio que se toma como base (siglo XV) o el que se transcribe exclusivamente. La actualización lingüística de las copias a lo largo de la Edad Media y de la difusión por medio de la imprenta manual (hasta el siglo XIX) es una realidad evidente para todos los que hemos tenido que enfrentarnos a una edición científica. Los sueños positivistas del siglo XIX de “reconstrucción lingüística” del original han quedado en eso: en pesadillas editoriales. Por otro lado, gracias a la distinción clara entre *texto* y *testimonio*, se podrían incluir dentro del proyecto las transcripciones de varios testimonios de un mismo texto, con nuevas posibilidades de herramientas informáticas (y de modificación de algunas existentes) para su uso científico. Y este punto puede relacionarse con el anterior, para permitir también la introducción de transcripciones paleográficas (en especial de documentos) que permitieran estudios grafemáticos y de fonética y fonología histórica, por sólo poner unos ejemplos.

El CORDE, como ya se ha indicado, es un proyecto científico en marcha, en el que se lleva trabajando desde 1994, pero que todavía le queda mucho esfuerzo por delante. ¿Qué es lo que se ha conseguido hasta ahora? Sus (casi) 300 millones de registros constituyen el mayor almacén de ediciones y de transcripciones que hasta ahora se había podido soñar; el mayor banco de datos textuales de la historia de la filología hispánica. Pero al primarse –en gran medida– la cantidad, el número de bites que conforman este enorme almacén, se ha desatendido –o al menos, así se percibe como usuario– el aspecto filológico, lo que podrá corregirse –si así se considera oportuno– en los próximos años. ¿Acaso la tecnología informática no ofrece más posibilidades si sabemos aunarla con los conocimientos humanísticos, como nos enseña la *Informática Textual*, y no quedarnos simplemente en la acumulación de la *información*, que nunca se ha de confundir con *conocimiento*?



## 5. La buena hora: a modo de primeras conclusiones

Gabriel García Márquez tituló una de sus obras *La mala hora*, pero no me ha parecido buen título para unas conclusiones, unas primeras conclusiones sobre el reto que para la edición y estudio de los textos medievales ofrecen las nuevas tecnologías informáticas.

Es cierto que el pasado más próximo no nos ha dejado una buena cosecha: el silencio y el desprecio parecen ser las aptitudes más habituales por la mayoría de los filólogos; y los proyectos que se han acercado a este ámbito ofrecen resultados muy criticables, dado que se han realizado de espaldas a la filología.

Pero estas limitaciones pueden convertirse en el punto de partida de nuestro trabajo futuro como filólogos, que no ha de caer en los mismos errores, ni esperar a que pasen cuatrocientos años para aceptar que la tierra gira alrededor del sol y darle razón a Galileo. De este modo, y como primeras conclusiones se ofrecen los siguientes cuatro puntos, cuatro columnas sobre los que alzar el edificio de la *Informática Humanística* en el ámbito hispánico:

- a) Aceptación de que la informática se ha convertido en una tecnología que no puede estar ausente de los estudios humanísticos.
- b) La necesidad de dar a conocer en el ámbito humanístico las herramientas informáticas que pueden ser útiles para nuestros propósitos, más allá de las habituales de comunicación, como el correo electrónico.
- c) La reflexión y creación de modelos hipertextuales para la difusión de los textos medievales en este nuevo medio que es Internet, teniendo en cuenta que es sólo un medio más, que en absoluto se ha de plantear como sustitución del medio analógico del libro, en su formato más tradicional.
- d) La necesidad de conocer la problemática textual en los proyectos informáticos que tienen al texto como objeto; la necesidad de devolver a la filología una voz en estas nuevas posibilidades que se nos ofrecen en el siglo XXI, tanto de trabajo, como de análisis como de edición.

**BIBLIOGRAFÍA / SITOGRAFÍA**

- BERNERS-LEE, TIM, *Tejiendo la Red*, Madrid, Siglo XXI de España Editores, 2001.
- CASTELLS, MANUEL, *La Galaxia Internet. Reflexiones sobre Internet, empresa y sociedad*, Madrid, Areté, 2001.
- CODINA, LLUIS, *El libro digital y la WWW*, Madrid, Tauro Ediciones, 2000.
- GIGLIOZZI G., *Il testo e il computer*, Mondadori, Milano 1997.
- , *La fondazione dell'informatica applicata al testo letterario*, a cura di Raul Mordenti, número especial de la serie "Testo e Senso", Editrice Universitaria di Roma, 2002.
- Informatica umanistica* (portal en Internet dedicado a esta disciplina: [http://www.griseldaonline.it/informatica/3sommario\\_index.htm](http://www.griseldaonline.it/informatica/3sommario_index.htm): 16 de septiembre 2004)
- LANDOW, George P., *Hipertexto. La convergencia de la teoría crítica contemporánea y la tecnología*, Barcelona, Paidós, 1995.
- LUCÍA MEGÍAS, José Manuel, "Editar en Internet (che quanto piace il mondo è breve sogno)", *Incipit*, XVIII (1998), pp. 1-40.
- , *Filología Románica en Internet. I. Los textos*, Madrid, Castalia, 2002.
- , "La "Informática Humanística": notas volanderas desde el ámbito hispánico", *Incipit*, XXIII (2003), pp. 91-114.
- , "Informática textual: nuevos retos para la edición y difusión de los textos (bibliotecas virtuales y bancos de datos textuales)", Seminario Menéndez Pidal, Universidad Complutense de Madrid, en prensa.
- ORLANDI Tito, *Informatica umanistica*, Nuova Italia Scientifica, Roma 1990.
- *Per l'Informatica nella Facoltà di Lettere*, Roma, Bulzoni, 1990.
- "Proposta: Informatica applicata alle discipline umanistiche (ovvero: Informatica umanistica)", *Griseldaonline* (<http://www.griseldaonline.it/informatica/index.htm>): 16 de septiembre 2004.
- PAJARES TOSCA, Susana, *Literatura digital. El paradigma hipertextual*, Cáceres, Universidad de Extremadura, 2004.
- RICCIARDI, Mario (cur.), *Scrivere comunicare apprendere con le nuove tecnologie* (I nuovi strumenti del sapere umanistico), Torino, Bollati Boringhieri, 1995.