

# El texto digital y la disyuntiva de las humanidades digitales

Ernesto Priani<sup>1</sup>

Recibido: 2015-08-29  
Enviado a pares: 2015-08-29

Aprobado por pares: 2015-09-29  
Aceptado: 2015-10-01

DOI: 10.5294/pacla.2015.18.4.11

## Para citar este artículo / To reference this article / Para citar este artigo

Priani, E. (Diciembre de 2015) El texto digital y la disyuntiva de las humanidades digitales. *Palabra Clave*, 18(4), 1215-1234. DOI: 10.5294/pacla.2015.18.4.11

## Resumen

A partir del hecho de que las humanidades se estudian dentro de una ecología mixta entre lo impreso y lo digital, y en donde este último está tendiendo a ser predominante, el artículo se pregunta qué tanto el cambio de ámbito conduce a las humanidades hacia el umbral de un nuevo paradigma, o si los conceptos, métodos e ideas que han constituido tradicionalmente a las humanidades continúan vigentes aún en este nuevo entorno. Para responder a esta pregunta, el artículo se propone abordar el tema del texto y sus posibles variaciones prácticas y teóricas a partir de su digitalización. El texto ocupa un lugar predominante dentro de las humanidades porque es, al mismo tiempo, objeto primario de estudio, producción de conocimiento y disseminación. El artículo sostiene que aun cuando la digitalización nos obliga a redefinir nuestra noción de 'texto' porque se ha modificado su naturaleza y se ha abierto una intensa discusión sobre lo que este es, no se ha roto ni teórica ni prácticamente con el paradigma de lo impreso. Las teorías sobre el texto digital desarrolladas por Dino Buzzetti, Manfred Thaller, Jerom McGann y Paul Caton, así como la concepción misma de la codificación SGML/XML/TEI, que serán objeto de análisis de este artículo, preservan residuos de la concepción tradicional sobre el texto, que evidencian

---

<sup>1</sup> Universidad Nacional Autónoma de México, México. epriani@gmail.com

cuán poco las humanidades se han desprendido de nociones provenientes de la tecnología de lo impreso.

### **Palabras clave**

Texto, humanidades digitales, digitalización, ecología de medios (Fuente: Tesouro de la Unesco).

## The Digital Text and the Choice of Digital Humanities

### **Abstract**

A part from the fact that the humanities are studied within a mixed ecology between print and digital, and where the latter is tending to prevail, the article questions how much the change of scope leads the humanities to the threshold of a new paradigm, or whether the concepts, methods and ideas that have traditionally been the humanities are still valid even in this new environment. To answer this question, the article attempts to address the topic of the text and its possible practical and theoretical variations from digitization. The text plays a major role within the humanities because it is at the same time, the primary object of study, knowledge production and dissemination. The article argues that although digitalization requires us to redefine our notion of 'text' because it has changed its nature and has opened an intense discussion about what this is, is not broken either theoretically or practically with the paradigm of what is printed. Theories of digital text developed by Dino Buzzetti, Manfred Thaller, Jerom McGann and Paul Caton as well as the very concept of coding SGML / XML / TEI, which will be analyzed in this article, preserve residues on the traditional concept of the text, which show how little the humanities have become detached from notions of print technology.

### **Keywords**

Text, digital humanities, digitization, media ecology (Source: Unesco Thesaurus).

# O texto digital e a disjuntiva das humanidades digitais

## Resumo

A partir do fato de que as humanidades são estudadas dentro de uma ecologia híbrida entre o impresso e o digital, na qual este último tende a predominar, este artigo se questiona sobre o quanto a mudança de âmbito conduz as humanidades ao limiar de um novo paradigma, ou se os conceitos, métodos e ideias que constituíram tradicionalmente as humanidades continuam válidas ainda nesse novo ambiente. Para responder a essa pergunta, o artigo se propõe abordar o tema do texto e suas possíveis variações práticas e teóricas a partir da digitalização. O texto ocupa um lugar predominante dentro das humanidades porque é, ao mesmo tempo, objeto primário de estudo, produção do conhecimento e disseminação. Além disso, sustenta que, ainda quando a digitalização nos obriga a redefinir nossa noção de texto, porque sua natureza foi modificada e aberta uma intensa discussão sobre o que este é, não se rompeu, nem teórica nem praticamente, com o paradigma do impresso. As teorias sobre o texto digital desenvolvidas por Dino Buzzetti, Manfred Thaller, Jerom McGann e Paul Caton, bem como a concepção da codificação SGML/XML/TEI, que serão objeto de análise deste artigo, preservam resíduos da concepção tradicional sobre o texto, o que evidencia que as humanidades têm se desprendido pouco das noções provenientes da tecnologia do impresso.

## Palavras-chave

Texto, humanidades digitais, digitalização, ecologia das mídias (Fonte: Tesouro da Unesco).

El presente artículo parte de la premisa de que las humanidades experimentan una transformación, producto de una alteración del entorno tecnológico dentro del cual tienen lugar. Si bien es cierto que el proceso está en curso y las humanidades se estudian todavía en una ecología mixta entre lo impreso y lo digital (Eggert, 2005), en la medida en que esta última va convirtiéndose en el ámbito predominante para la producción, el estudio y la diseminación de las humanidades, cabe la pregunta de qué tanto el cambio de ámbito conduce a las humanidades hacia el umbral de un nuevo paradigma, o si los conceptos, métodos e ideas que han constituido tradicionalmente a las humanidades continúan vigentes aún en este nuevo entorno. Para responder así sea parcialmente a esta pregunta, me he propuesto abordar el tema del texto y sus posibles variaciones prácticas y teóricas a partir de su digitalización. Texto es, sin lugar a dudas, uno de los conceptos más trascendentes para el conjunto de las humanidades que, al menos en principio, cambia de sentido con la nueva ecología digital. De hecho, la cuestión de la textualidad es uno de los problemas en que se ha centrado el debate dentro de las llamadas humanidades digitales (Eggert, 2005, Caton, 2013, Buzzetti y Thaller, 2012), ese espacio de aplicación y reflexión académica sobre el uso de métodos y sistemas computacionales en las humanidades. La razón resulta obvia: el texto ocupa un lugar predominante dentro de las humanidades porque es, al mismo tiempo, objeto primario de estudio, producción de conocimiento y diseminación. Su digitalización parece modificar radicalmente la concepción del texto porque introduce diferencias importantes en su producción, manipulación y conservación (Hillesund, 2005), y en esa medida abre nuevas opciones para su estudio y producción dentro de las humanidades. Pero ¿qué tanto la digitalización del texto ha significado una modificación de su concepción clásica, como un ideal que precede a cualquiera de sus formas, habiendo sido concebido únicamente por el autor (Cerquiglini, 1998)? ¿Qué tanto significa una renovación de medios y métodos de estudio? En suma, ¿qué tan lejos realmente estamos del texto “autónomo y cerrado”, como lo definirían Todorov y Ducroct y de las formas de su análisis (Culler, 2007)? Lo que voy a argumentar a continuación es que, aun cuando la digitalización nos obliga a redefinir nuestra noción de ‘texto’ porque se ha modificado su naturaleza y se ha abierto una intensa discusión sobre lo que este es, no se ha roto ni teórica ni práctica-

mente con el paradigma de lo impreso. Con ese fin, me propongo mostrar que tanto las teorías sobre el texto digital desarrolladas por Dino Buzzetti, Manfred Thaller, Jerom McGann y Paul Caton, así como la concepción misma de la codificación SGML/XML/TEI, que serán objeto de análisis de este artículo, preservan residuos de la concepción tradicional sobre el texto, que evidencian cuan poco las humanidades se han desprendido de nociones provenientes de la tecnología de lo impreso. En suma, sostendré que el camino hacia una ecología de digital no significa, todavía, una revolución dentro las humanidades, sino solo un paso dentro de un proceso evolutivo.

## ¿Un texto, muchos textos?

El texto digital es casi tan antiguo como el cómputo. En un artículo clásico, Bush (1945) imaginó una máquina a la que denominó Memex, que presentaba como un sistema en el que sería posible almacenar libros y archivos, y dentro del cual se pudiera mecanizar su consulta, de manera rápida y flexible. Memex permitiría, además, hacer anotaciones a los textos y relacionar estos entre sí, a partir de las propias anotaciones o las decisiones del lector, de modo que la organización dependiera de los intereses del usuario y no del catalogador.

Para el teórico del hipertexto Landaw (1991), Bush reconocía en 1945 la necesidad de un texto virtual en lugar de un texto físico, para que las relaciones entre diversas obras estuvieran dadas por su lectura y no por su clasificación. Con ello, al menos de manera teórica, se abrió la puerta a una nueva concepción del texto que no llegaría a realizarse sino hasta unas tres décadas después, con la aparición de los proyectos pioneros de digitalización de textos, como el Proyecto Gutenberg en 1971 (Hart, 1992).

A diferencia de otros proyectos en la lingüística y de la estilística, que fueron de los primeros en utilizar el procesamiento computacional para obtener resultados a partir del análisis de textos (Buzzetti y Thaller, 2012), la producción de ediciones digitales inauguró un nuevo campo problemático en relación con la textualidad porque se tuvo que resolver una cuestión crucial para poder llevarse a cabo: ¿cómo se presenta el texto en la pantalla de una computadora?

La decisión tomada por el Proyecto Gutenberg fue hacerlo en Plain Vanilla ASCII (Hart, 1992), es decir, utilizando una codificación compuesta por un set mínimo de caracteres definidos numéricamente, en un archivo que no ofrecía información más allá del texto mismo, para que pudiera ser leído en cualquier sistema (figura 1).

### Figura 1. Los dos primeros versos de *Amor condusse noi ad una norte*, de José Villaurrutia en código ASCII

Amar es una angustia, una pregunta, una suspensa y luminosa duda
065 109 097 114 032 101 115 032 117 110 097 032 097 110 103 117 115 116 105 097 044 032 117 110 097 032 112 114 101 103 117 110 116 097 044 013 010 117 110 097 032 115 117 115 112 101 110 115 097 032 121 032 108 117 109 105 110 111 115 097 032 100 117 100 097 059

El archivo no contenía ningún tipo de metadatos, pero tampoco características tipográficas, como subrayados, negritas, páginas, encabezados o subtítulos; algunas de esas variantes, como los títulos, se representaban utilizando mayúsculas. Vistos a la distancia, los primeros textos digitales presentaban un texto apenas suficiente para ser comprensible para un lector. En ellos podemos observar, por primera vez, la diferencia entre los datos almacenados en la computadora y el texto que aparecía en la pantalla. La diferencia estaba dada por la codificación ASCII que traducía los datos elementales del texto, como las letras, los espacios entre palabras, la puntuación, a un código numérico, para que, desde ahí, la computadora lo interpretara para su posterior representación en el espacio de una pantalla (figura 1). Para Hillesund (2005) este es el cambio más significativo de la digitalización respecto del proceso de producción textual en lo impreso:

The important and extremely consequential point of the digital shift is that the storing of the text is separated from its visual representation. In a digital text cycle, units of text are stored as bit patterns (usually in the codes of ASCII or Unicode) in a file. When writing is finished, the whole text is stored as a collection of bit patterns in a mass storage system, often a hard disk, completely separate from any representation of the text.

Esta separación entre la forma de almacenar el texto y presentarlo ha sido llevada al campo teórico. Como han señalado, entre otros, Buzzetti y

McGann (2006), para la computadora el texto es “información codificada como caracteres o secuencia de caracteres”. Para que tal secuencia adquiriera sentido para un lector humano, es necesario que incorpore códigos específicos (ASCII, XML), que instruyan a la computadora sobre el modo de interpretar esos caracteres. Ese código, que acaba por ser parte del texto mismo, permite distinguir entre el texto (privado de código) y el que toma forma en la pantalla.

Pero la diferencia entre el texto y su representación ha sido señalada por Buzzetti y Rehbein (1997) también para el caso de libro, donde la codificación estaría dada por las características materiales, tipográficas, etcétera:

It is true that “most of us almost automatically equate texts with printed books”, but we should be wary of confusing “the form of representation”, or the properties of the book, “with the form of what is to be represented” (Buzzetti y Rehbein, 1997).

Una inferencia que no deja de ser discutible porque equiparar las características físicas del libro con la codificación implica darle al texto impreso las características del texto digital, lo cual cuestiona enfáticamente, por ejemplo, Hillesund (2005).

En todo caso, la ruptura de la identidad entre el texto y su representación, que los métodos de almacenamiento digital vuelven paradójicamente tangible, conduce a una pregunta que, desde las llamadas humanidades digitales, ha intentado ser respondida: ¿qué es texto?

## Dos teorías acerca del texto

De las aproximaciones teóricas al texto en las humanidades digitales, han destacado las propuestas por Dino Buzzetti en colaboración con distintos autores, y la que de manera más reciente ha propuesto Caton (2013). Ambas sintetizan, a mi juicio, los ejes en que el debate se ha venido desarrollando en los últimos años, y muestran las ventajas y los límites teóricos de las propuestas.

Para Buzzetti y Rehnein (1997), el texto no es propiamente definido, pero señalan: “A textual representation is by necessity concrete and needs a physical support. Text, on the contrary, is abstract and may be conceived as the invariant content of all its material representations”.

*Abstracto* es un término cauteloso para caracterizar al texto, pues lo describe solo a partir de la ausencia de cualidades concretas, reservadas todas ellas a su representación. Pero abstracto supone, también para los autores, que es inexistente porque “a text comes into existence only through the representation of its content”.

Con el concepto de ‘texto’ se estaría enfrentando aquí a los mismos problemas que toda teoría, en la que se postula un modelo que subyace a la existencia de expresiones concretas. Al describirlo como abstracto, invariante e inexistente, Buzzetti y Rehnein (1997) otorgan al texto un carácter semejante a lo *noúmenico* kantiano, pues entenderíamos por texto algo que no puede ser objeto de nuestra comprensión y lo suponemos “sólo como una abstracción de sus posibilidades sensibles” (Kant, 1972/2009).

La renuncia a la idealidad del texto, propio de la filología clásica, para postular su abstracción ciertamente coloca el valor principal, la existencia, en las formas concretas, a diferencia de hacerlo en la forma puramente ideal. Pero esta fórmula, aunque constituye una variante, no rompe por completo con la idea de que el texto posee una identidad última, que es distinguible de sus expresiones individuales. Más adelante volveré sobre esta cuestión y las limitaciones que conlleva dentro de la concepción de las humanidades.

Por su parte, Caton (2013) define \*texto\*<sup>2</sup> a partir de tres características que considera fundamentales. \*texto\* es “la representación escrita del lenguaje” que “muestra el intento de comunicar algo” y que “está completo dentro de su contexto” (*being complete within its context*). Es importante subrayar que las dos primeras características producen “an uncountable \*text\*<sup>\*</sup>”. *Uncountable* es un término de medida y es utilizado por Caton para hacer hincapié en que \*texto\*<sup>\*</sup>, en cuanto es definido por las dos

---

2 Esta es la notación utilizada por él, es para acotar el uso del término al campo de las humanidades digitales.

primeras características, no es conmensurable. Por lo mismo, no es discernible, sino salvo cuando está completo en relación con un contexto.

Este \*text\* inconmensurable es para él la materia textual. En sus palabras: “I suggest that the combination of characteristics one and two gives us \*text\* in the mass noun sense: \*text\* as material, as stuff”.

Examinemos con más cuidado esta afirmación, tratando de entender de dónde proviene el carácter material, de cosa, que le otorga al texto. Caton es muy enfático al señalar que no intenta definir texto en general, sino solo el uso del término en el contexto de las humanidades digitales y, aún de manera mucho más restrictiva, su uso en el contexto de la codificación XML de las *Guidelines for Electronic Text Encoding and Interchange* (TEI) (Caton, 2013).

Tomando en consideración estas advertencias, resulta más claro que Caton concibe \*texto\* como aquello dado que es una representación escrita del lenguaje que intenta comunicar algo, a lo cual el contexto dará un sentido. Es difícil no pensar, incluso por los ejemplos ofrecidos por él, que se refiere con “an uncountable \*text\*” a entidades preexistentes a la codificación: libros, carteles, telegramas, panfletos. De ahí el sentido material, de objeto que le confiere al texto. Pues contexto, como él aclara hacia el final del artículo, puede referirse a la primera edición de una obra o a una cita en un cartel, que se refiere a su vez a un discurso más extenso.

Sorprende, sin embargo, que ese \*texto\* —el que está dado por las dos primeras características— sea precisamente *uncountable*, cuando es concreto. Precisamente porque supone que no tiene límites establecidos de antemano. En este sentido, su concreción en un libro o en un cartel no significa necesariamente que sea conmensurable. La definición de \*texto\* con las tres características integradas solo se produciría una vez que se vincule a un contexto determinado que, entiende, es definido por el editor digital del texto.

Through questioning the notion of being complete-within-contextual-bounds I concluded that being \*a text\* is a status we give some

quantity of \*text\* (mass) in a particular context and at our choosing. Use of the definite article implies that the context is understood. Normally, the context will be an abstraction that we always think of as bounded. Thus we speak of, for example, “the text of the first edition”, where the abstract entity signified by “the first edition” supplies the conceptual boundary (Caton, 2013).

En este sentido, para Caton, la edición de las *900 tesis* de Pico della Mirandola hecha por Eucharius Silber en Roma en 1486, por ejemplo, podría ser, al mismo tiempo \*text\* o definir el contexto de un \*a text\*, pues el editor podría ver en ella una unidad rota con el *Discurso sobre la dignidad del hombre* y la *Apología*, de Pico, e integrar las tres obras, apelando al contexto en que se produce la disputa romana de Pico. En ese sentido, las tesis serían \*text\* hasta que se completara con la adición de las otras dos obras. Pero la misma edición de Eucharius Silber podría ser la base contextual de un \*a text\* tomada de manera singular, si el editor decide que esa edición es suficiente para entender conceptualmente la obra. En suma, el editor decide cuándo una cierta “masa textual” es un texto, cerrado y completo, y cuándo no.

A pesar del mayor realismo dentro del cual intenta moverse Caton, tampoco renuncia del todo a la idea de que el texto tiene una identidad más allá de sus representaciones concretas. Que esta identidad sea tratada como material es solo relativa, pues en el fondo es inconmensurable y, por tanto, indefinida.

Lejos de oponerse entre sí las definiciones analizadas hasta aquí, ambas resultan ser muy semejantes. Supone que hay que pensar el texto de dos maneras distintas. Por un lado, una que podríamos llamar de origen, que caracteriza al texto por su indeterminación y, por otro, una donde texto es determinado, ya sea por la representación, ya sea por la decisión del editor. Ambas separan el contenido del medio que lo transmite, y suponen que la codificación es capaz de restituir el texto original en una infinidad de variantes.

Esta semejanza en la estructura de la reflexión sobre el texto, a pesar de una escalonada renuncia a la idealidad, lleva a pensar que la discusión sobre la naturaleza del texto no se limita a identificar una posición “más realista”,

como la de Caton (2013), o “menos realista”, como la de Buzzetti y Rehbein (1997), y posicionarse entre ellas, sino en discutir lo que está a la base de ambas, y qué comparten. Como señalan Fiormonte, Geri y Silvi (2006):

[...] l'esigenza di ricostruzione del testo ereditata dalle filologie classica e medievale si è sposata felicemente con l'idea di modello astratto (Buzzetti 2002), e con la promessa che a partire da un unico “meta-testo” (il documento codificato) fosse possibile ricavare diverse edizioni su diversi supporti.

La crítica puede dirigirse con las mismas palabras a Caton, quien independientemente de que no imagine el texto como una entidad abstracta, sí la piensa como inconmensurable, lo cual plantea un problema semejante. En realidad, la cuestión aquí recae en que esta idea de la identidad última del texto con su forma indeterminada no es sino un residuo, menos idealista, más fenomenológico, de la idea clásica de texto. No podemos negar que en estas propuestas teóricas haya un movimiento, un redimensionamiento del concepto a la luz del nuevo entorno digital, pero este no es tan amplio como para constituir una idea que se haya desprendido realmente de lo impreso y de los principios de la filología clásica formada en él.

Este desplazamiento no es, sin embargo, privativo de la teoría. Lo encontramos en el corazón mismo de la concepción del texto para proponer la forma y estructura de su codificación, con la pretensión de construir una representación que pudiera incluir la totalidad de sus variantes.

## ¿Representarlo todo?

Para Buzzetti y Robhen (1997), el marcado de texto con una codificación XML es entendido como “all the information in a document other than the ‘contents’ of the document itself, viewed as a stream of characters”. En este sentido,

[...] any information present about formatting (margins, font shifts, page breaks) is given by markup. All structural information, all relationships among texts, apparatus, and notes, is given by markup. All analytic or interpretive information to be included in an electronic text is by definition expressed as markup.

Como se desprende de esta cita, el proceso de codificación con el marcado del texto agrega a este un espectro enorme de información, con la cual, al menos idealmente, podría formarse un único metatexto que representarse todas las variantes posibles de un mismo texto, como esperaba Cerquiglini (1999) que podría hacerse con la inclusión del texto digital en la filología.

Pero el problema es que la codificación tiene límites que hacen imposible, incluso no deseable, cumplir con esta promesa. Las decisiones tomadas para representar el texto con la codificación SGML/XML/TEI son una extensión de la concepción de texto impreso (Hillesund, 2005) y, en esa medida, fue concebida como una extensión instrumental para los estudios humanísticos.

El modelo para describir lo que es un texto responde, desde la década de 1980, con el establecimiento del SGML (Standard Generalized Markup Language) como modelo de codificación, a una serie de ideas que establecen los límites que aún tiene este sistema. En síntesis, estas ideas pueden ser enunciadas de la siguiente forma:

1. El texto tiene una fuente única (Hillesund, 2005).
2. Su estructura es anidada (DeRose et al., 1997).
3. No se define como un modelo de datos (Renear, Dubin y Sperberg-McQueen, 2002).
4. Su sentido tiene que ser definido más allá de la codificación (Dubin y Birnbaum, 2004).

Como se apuntaba apenas unos párrafos arriba, la promesa principal de la digitalización era la posibilidad de representar todas las variantes posibles de un texto. Paradójicamente, el modelo SGML/XML/TEI trata de lograr esa promesa a partir de mantener el principio de que el texto tiene una sola fuente y que, por tanto, puede ser separado de la forma en que aparece, sin que esto afecte al texto. El presupuesto es claro: la única manera de ofrecer todas las variantes de un texto es que este no dependa, de ninguna manera, de su expresión concreta, de manera que todas sus formas concretas puedan ser integradas finalmente en un solo metatexto.

Para Hillesund (2005) esta idea descansa “on the false supposition that any kind of written or verbal content can be presented at will in any medium and for whatever purpose”. Igualmente falsa es la idea de que todo texto puede ser asimilado a un texto único como resultado del proceso de su codificación no solo por sus inconvenientes prácticos (Fiormonte, Geri y Silvi, 2006), sino también teóricos, al plegar la multiplicidad de las formas en que aparece la escritura al espacio uniforme del texto codificado. Además, las propias limitaciones del modelo SGML/XML/TEI mostrarán las dificultades para cumplir con este ideal.

Desde su concepción en la década de 1980, el modelo de representación textual SGML/XML/TEI responde a la idea expresada en la tesis de DeRose et al. (1997) de que el texto es “essentially composed of nesting objects such as chapters, sections, paragraphs, extracts, lists, and so on” (Renear, Mylonas y Durand, 1993).

En la práctica, este modelo implica dos cosas: 1) que se renunció a otros modelos posibles de representación, como pueden ser una estructura de objetos diseñados, matrices de píxeles u otras utilizadas en el cómputo (Renear, Dubin y Sperberg-McQueen, 2002), incluso la estructura de bases de datos (Schloen y Scholen, 2014), y 2) que, al optar por una estructura de jerarquías de objetos anidados para representar el texto —una página forma parte de un capítulo, que forma parte de un libro, etcétera—, se dificulta presentar un mismo texto (o una fracción de él) bajo dos características diferentes. Por ejemplo, codificar las características tipográficas de una palabra y su función como título o diferenciar el nombre propio de la referencia a una persona, lo cual se conoce como el problema del *overlapping* (Renear et al., 2002; Schloen y Scholen, 2014, Eggert, 2005, Buzzetti y Thaller, 2012).

El *overlapping* representa una dificultad práctica porque obliga a representar cualquier aspecto del texto como una jerarquía ordenada, o bien como una subjerarquía ordenada (Renear, Mylonas y Durand, 1993), eliminando las posibilidades de representar, por ejemplo, “estructuras de información, como las variantes abiertas, que violan el principio de correspondencia” (Fiormonte y Martiradonna, 2010).

Incluso, desde un punto estrictamente pragmático, lo que hace que el lenguaje de marcado sea un éxito es su límite, porque “el grado de fedeltà alla fonte originaria è proporzionale al grado di astrazione che il linguaggio rende possibile, più aumenta la complessità del documento, più complesso e verboso il markup e maggiori le difficoltà de visualizzazione e dunque di accesso a esso (Fiormonte, Geri y Silvi, 2006).

Las reglas definidas SGML/XML/TEI, además, ofrecen solo un vocabulario y una sintaxis, pero no una semántica para ese vocabulario (Renear et al., 2002). En este sentido, como señalan Buzzetti y Thaller (2012), “for embedded markup provides a data structure, but it does not beget a suitable data model”, en la medida en que no hay “one-to-one correspondence between the logical form of a phrase and the structure of its semantic content”.

La conciencia principal de esto, al mismo tiempo una de las mayores fortalezas del modelo, es que cualquiera de las marcas (tómese por ejemplo, <title>, <place>, <name>) carece de un sentido definido y de unas reglas de uso explícitamente establecidas, lo cual le otorga, por un lado, una flexibilidad por la que cada investigador puede determinar cuáles son las formas en que esas marcas se utilizarán y qué representan, pero, al mismo tiempo, como señalan Dubin y Birnabaum (2004), significa que la información central para comprender el marcado no está presente en el marcado mismo, lo cual limita y dificulta muchas automatizaciones y procesamientos al no proveer una forma realmente abstracta del texto. De hecho, un mismo marcado puede tener diferentes sentidos en diferentes contextos, puede comunicar el mismo sentido de formas diferentes, autorizar inferencias en diferentes dominios semánticos simultáneamente, o bien ofrecer evidencia de supuestos sentidos que no pueden ser interpretados completamente basados solo en el marcado (Dubin y Birnabaum, 2004), lo cual problematiza mucho la creación de reglas semánticas que deben recurrir a información más allá del marcado para establecerse. Es decir, la codificación no contiene todos los elementos para su comprensión. De ahí la necesidad de buenas prácticas, como la definición del uso de las marcas y su sentido, para poder tener una comprensión total de las intenciones de la codificación.

Pero esta ambigüedad inherente al marcado de texto tiene otras consecuencias, a mi juicio, las más interesantes por sus posibilidades, pero también por las dificultades que conlleva.

La codificación digital, debe señalarse,

[...] no es una simple “grabación” de un documento en otro soporte, sino un proceso hermenéutico, semiótico (en el momento en que un texto se transcribe a través de la elección y la utilización de marcadores, el texto produce a su vez significado) y pragmático (los marcadores pueden representar acciones pero también crearlas (Fiormonte y Martiradonna, 2010).

La ausencia de una semántica de la codificación permite justamente que el ejercicio hermenéutico tenga lugar durante el proceso de codificación y quede reflejado precisamente en la elección y aplicación de las marcas. En ese sentido, las marcas son al mismo tiempo la ejecución del proceso de conocimiento y un registro de este. Uno de los aspectos más interesantes del marcado es que termina por mostrar solo cierta lectura del texto, la que realizan quienes finalmente llevan a cabo el marcado. Pero el hecho de que muestre una lectura constituye una limitación, pues cada marcado estaría imposibilitado de contener tanto estructural como analíticamente más que una representación parcial de las posibilidades del texto.

De ahí que Fiormonte y Martiradonna (2010) señalen:

Resultará imposible, pero también poco deseable, representar todas las dimensiones textuales, ya que si variamos la atención del producto al proceso de escritura nos damos cuenta de que no existe un Texto sino muchos textos, tantos como son los mecanismos de escritura, los recorridos intertextuales y las propuestas de reconstrucción.

## Conclusiones

El SGML/XML/TEI se construyó teniendo en mente el papel. Presupuestos básicos, como el texto único, la descripción del texto como un orden jerárquico (que tiene el libro y sus subdivisiones como modelo), la renuncia a intentar crear un modelo completamente abstracto, dejar las reglas semánticas

fuera de la codificación, permiten concluir que en su origen el modelo de codificación trasladaba ciertas ideas del ámbito de lo impreso a lo digital.

Este tránsito no puede ser tomado como una repetición. No persigue solo ni de manera central la reconstrucción de lo impreso en lo digital. Al contrario, se propone una tarea mucho más ambiciosa: la generación de un metatexto en que pudieran ser incluidas todas las variantes del texto, siempre en el entendido de que al final este es uno. En este sentido, el XML/TEI fue diseñado como un instrumento para llevar a cabo una instrumentalización de un cierto modelo filológico que, como se ha señalado, se refleja también en las posiciones teóricas más representativas sobre el texto en las humanidades digitales.

Estas posturas, como se señaló arriba, identifican el texto con una suerte de texto originario que tiene dos características dominantes: su carácter único (fuente de cualquier variante) y su ininteligibilidad (abstracción o inconmensurabilidad), mientras que a toda variante del texto le correspondería su determinación, sea esta entendida como existencia, sea como definición en un contexto, lo cual de hecho no anula, sino que de cierta forma traslada y actualiza el modelo filológico clásico. Estas teorías comparten también, al menos de manera hipotética, que la codificación del texto es capaz de representar cualquier variante que se presente.

Es cierto que, pese a las limitaciones teóricas y prácticas de la codificación, esta ha abierto la puerta a nuevas exploraciones sobre el texto que, al tiempo que ensanchan nuestro horizonte de conocimiento, ofrecen evidencia de las limitaciones del modelo. En este sentido, podemos concluir que la codificación del texto no representa una revolución, es decir, un rompimiento profundo y claro con las concepciones y prácticas del trabajo académico en un entorno de lo impreso, sino una evolución en el campo de las humanidades, de la misma forma en que para Hillesund (2005) el texto digital representa una evolución y no una revolución en la modificación de los ciclos textuales. Es decir, se trata de un cambio relativo que establece la modificaciones de prácticas e ideas teóricas, del que es posible esperar constituya la base de otros más profundos.

Por lo pronto, las propuestas hechas, entre otros, por Egger (2005) y Buzzetti y Taller (2012), de superar el problema de la codificación del texto mediante modelos de *stand-off-markup*, en el que se propone extraer el marcado de su imbricación con el texto para evitar el *overlapping*. Pero no solo superar este problema, la propuesta también busca restituir al texto digital su carácter, su condición digital, “sin esperar que represente por él mismo el texto impreso o manuscrito” (Egger, 2005). Buzzetti y Taller (2012), al comentar el proyecto BECHAMEL, apuntan que un sistema de *stand-off-markup* atendería también el problema de la separación entre el vocabulario y la sintaxis y las reglas semánticas que se presentan en la codificación tradicional. Un camino por el que se sigue trabajando. Existe también la propuesta más reciente de Schloen y Schloen (2014) de trasladar el texto digital a una base de datos, con la intención de abandonar “el paradigma del documento”, aprovechando que las bases de datos han alcanzado una sofisticación que permite representar en ellos estructuras documentales.

Estas perspectivas no solo ofrecen soluciones específicas a los problemas prácticos y teóricos planteados alrededor del texto digital, sino que señalan que los horizontes de búsqueda comienzan a orientarse más allá del entorno de lo impreso hacia lo propiamente digital.

En suma, las humanidades han entrado en un proceso de evolución para integrarse, de manera progresivamente más profunda, a la ecología de lo digital.

## Referencias

Bush, V. (1945). We may think. *The Atlantic Monthly*, 176(1), 101-108.

Buzzetti, D. y Thaller, M. (2012). Beyond embedded markup. *Journal of Computing in Higher Education*, 1(2), 3-2. Recuperado el 16 de junio de 2014 de [http://web.dfc.unibo.it/buzzetti/dbuzzetti/publicazioni/hamburg\\_dh2012\\_ed.pdf](http://web.dfc.unibo.it/buzzetti/dbuzzetti/publicazioni/hamburg_dh2012_ed.pdf)

Buzzetti, D. y McGann, J. (2006). Critical editing in a digital horizon. *Electronic Textual Editing*, 53-73. Recuperado el 16 de junio de 2014 de [http://www.tei-c.org/About/Archive\\_new/ETE/Preview/mcgann.xml](http://www.tei-c.org/About/Archive_new/ETE/Preview/mcgann.xml)

- Buzzetti, D. y Rehbein, M. (1997). Textual fluidity and digital editions. *Text Variety in the Witnesses of Medieval Texts*, 14-39. Recuperado el 2 de agosto de 2014 de <http://www.denkstaette.de/files/Buzzetti-Rehbein.pdf>
- Caton, P. (2013). On the term 'text' in digital humanities. *Literary and Linguistic Computing*, 28(2), 209-220.
- Cerquiglini, B. (1999). *In praise of the variant: a critical history of philology*. Baltimore: Johns Hopkins University Press.
- Culler, J. (2006). *The literary in theory*. Stanford: Stanford University Press.
- DeRose, S. J., Durand, D. G., Mylonas, E. y Renear, A. H. (1997). What is text, really? *Asterisk Journal of Computer Documentation*, 21(3), 1-24.
- Dubin, D. y Birnbaum, D. J. (2004, agosto). Interpretation beyond markup. En B. T Usdin (ed.), *Proceedings of Extreme Markup Languages*. Montreal, Quebec. Recuperado el 2 de agosto de 2014 de <https://www.ideals.illinois.edu/bitstream/handle/2142/11838/EML-2004Dubin01.pdf?sequence=2>
- Eggert, P. (2005). Text-encoding, theories of the text, and the 'work-site'. *Literary and Linguistic Computing*, 20(4), 425-435.
- Fiormonte, D. y Martiradonna, V. (2010). La representación digital de la génesis del texto: un caso de estudio. En *En el taller del escritor: génesis textual y edición de textos* (pp. 147-176). Bilbao: Universidad del País Vasco.
- Fiormonte, D., Geri, L. y Silvi, D. (2006). XML e i testi contemporanei: problemi di codifica, rappresentazione e analisi. Recuperado el 3 de septiembre de 2014 de [https://www.academia.edu/3493342/XML-TEI\\_and\\_contemporary\\_literary\\_texts\\_encoding\\_representation\\_and\\_analysis](https://www.academia.edu/3493342/XML-TEI_and_contemporary_literary_texts_encoding_representation_and_analysis)

- Hart, M. (1992). The history and philosophy of project Gutenberg. Recuperado 3 de septiembre de 2014 de [https://www.gutenberg.org/wiki/Gutenberg:The\\_History\\_and\\_Philosophy\\_of\\_Project\\_Gutenberg\\_by\\_Michael\\_Hart](https://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart)
- Hillesund, T. (2005). Digital text cycles: from medieval manuscripts to modern markup. *Journal of Digital Information*, 6(1). Recuperado el 4 de octubre de 2014 de <https://journals.tdl.org/jodi/index.php/jodi/article/view/62/65>
- Kant, E. (2009). *Critica de la razón pura* (trad. M. Caimi). México: Fondo de Cultura Económica.
- Landow, G. P. (1991). *HyperText: the convergence of contemporary critical theory and technology (parallax: re-visions of culture and society series)*. Baltimore: Johns Hopkins University Press.
- Renear, A., Dubin, D. y Sperberg-McQueen, C. M. (2002, noviembre). Towards a semantics for XML markup. En *Proceedings of the 2002 ACM symposium on Document engineering* (pp. 119-126). Nueva York: ACM.
- Renear, A. H., Mylonas, E. y Durand, D. (1993). Refining our notion of what text really is: The problem of overlapping hierarchies. En N. Ide y S. Hockey (eds.), *Research in Humanities Computer*. Oxford: Oxford University Press. Recuperado el 4 de octubre de 2014 de <https://www.ideals.illinois.edu/bitstream/handle/2142/9407/RefiningOurNotion.pdf?sequence=2>
- Schloen, D. y Schloen, S. (2014). Beyond Gutenberg: transcending the document paradigm in digital humanities. *Digital Humanities Quarterly*, 8(4). Recuperado el 4 de octubre de 2014 de <http://www.digitalhumanities.org/dhq/vol/8/4/000196/000196.html>

