

De XML a PDF Tipografía con T_EX

Javier Bezos

Presidente de CervanT_EX
Grupo de Usuarios Hispanohablantes de T_EX
presidente@cervantex.org
Compositor tipográfico
jbezos@wanadoo.es

Resumen Uno de los problemas de los sistemas de visualizado con HTML es el escaso control tipográfico que se tiene sobre el resultado final. Los archivos PDF, por el contrario, permiten un control total sobre el formato, pero ello plantea el nuevo reto de crear, a partir de XML, documentos PDF con la calidad tipográfica necesaria y adaptados a las pautas ortotipográficas que uno puede esperar en un libro. El sistema T_EX soluciona este problema, lo que permite que la presentación de los contenidos sea más familiar y por tanto más efectiva.

Creado hace 25 años por Donald E. Knuth, T_EX ha resultado ser una pieza esencial en los flujos de composición tipográfica automática debido a su estabilidad, su larga implantación en medios académicos, su libre acceso, la experiencia acumulada durante ese tiempo y, naturalmente, la calidad de los resultados.

La exposición estará dedicada a hacer un breve resumen de la historia de T_EX, la idea básica de su funcionamiento, la influencia otros sistemas, incluido XML y ejemplos de sus posibilidades.

1. Introducción

Esta exposición no busca resolver un problema; más bien al contrario, su objetivo es mostrar su existencia. José Martínez de Sousa [1998] lo expone del siguiente modo:

El problema radica en que la máquina es maravillosa y los programas que en ella se utilizan son asimismo maravillosos, pero el hombre que los maneja solo sabe, desde el punto de vista bibliológico y tipográfico, eso: manejar la máquina y los programas.

Menos rotundo, pero más preciso, es Yannis Haralambous cuando habla de la tipografía en su prefacio de *Digital Typography Using L^AT_EX* (Syropoulos [2003]):

[...] a craft that very few people actually have studied, some people have learned by themselves – mainly by actually making books – and most people are generally unaware of.¹

¹ «[...] un oficio que muy pocos han estudiado realmente, algunos han aprendido por sí mismos —sobre todo componiendo libros— y la mayoría en general desconoce.»

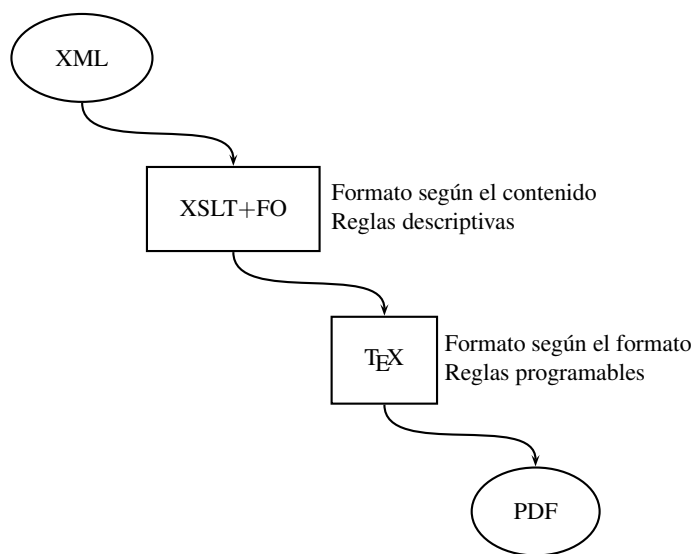


Figura 1. Esquema simplificado del paso de XML a PDF

Es decir, nos encontramos ante una disociación entre la herramienta y el oficio, de tal forma que la herramienta, es decir, los sistemas informáticos, han pasado a ser el oficio, y que el auténtico oficio, la tipografía, ha quedado relegado a un segundo plano o, simplemente, ha desaparecido. Toda innovación se analiza desde el punto de vista técnico de los medios, pero a menudo se pierde de vista el fin que se busca con ellos. Es preciso recordar, y es lo que en el fondo se pretende aquí, que esos «maravillosos» medios, que sin duda lo son, tienen que medirse también por los resultados tipográficos.

2. La transformación de XML a PDF

El mecanismo seguido normalmente para la transformación de XML a PDF está esquematizado en la figura 1, donde el proceso de conversión ha quedado sucintamente reducido a dos pasos esenciales: la transformación de XML a una forma que un programa de composición pueda entender, y la creación del archivo PDF por ese programa. El primer paso está indicado como XSLT+FO, pero igualmente podría ser DSSSL o incluso `xmLtex` o `dbLatex`;² el método concreto es irrelevante para esta discusión.

También hay varias posibilidades para el segundo paso: Java y FOP³ no son particularmente sofisticados, y de `TeX` se suele hacer un uso más bien mecánico. Sin embargo, es posible, y conveniente, sacar partido de las posibilidades de `TeX` con un uso más creativo. En particular:

² <http://sourceforge.net/projects/dblatex>

³ <http://xml.apache.org/fop>

Normas tipográficas. La influencia de los programas de maquetación se ha dejado sentir en los mecanismos de formato usados en XML. Más preocupados por los aspectos macrotipográficos, los detalles microtipográficos —según unas reglas implícitas que normalmente caen dentro de las tareas del corrector tipográfico— son a menudo desconocidos.

Formato en función del formato. Las transformaciones que hay que aplicar no sólo pueden depender del contenido, sino también del formato. Por ejemplo, podemos exigir que los cuadros ocupen toda la medida del texto, y si no hubiera espacio suficiente, que sea la medida más dos centímetros. Ello no se opone a las transformaciones según el contenido, sino que más bien se introduce un nivel adicional.

Reglas programables. Hay reglas que quedan fuera de cualquier intento de sistematización. Son muy específicas de las situaciones concretas o de la naturaleza de la obra y es imposible crear parámetros que cubran absolutamente todas las posibilidades. Disponer de un entorno de composición programable puede resultar esencial en estos casos.

El reto que plantea el primero de estos puntos no es pequeño. Programas como QuarkXPress, PageMaker e InDesign pueden importar XML, pero los detalles tipográficos siguen quedando en manos de los que lo usan, en un proceso que en buena medida sigue siendo artesanal. Una de las principales virtudes de XML es que se puede omitir por completo cualquier detalle del aspecto, pero en un proceso automatizado eso implica que no podemos intervenir manualmente para controlar los detalles tipográficos. El objetivo es, por tanto, mostrar cómo obtener un archivo PDF tipográficamente correcto a partir de un documento XML que carezca por completo de toda indicación tipográfica.

3. Breve historia de T_EX

La historia de T_EX comienza a mediados de los años setenta, cuando la American Mathematical Society (AMS) tenía serias dificultades para componer sus publicaciones. Los sistemas de fotocomposición podían realizar la tarea, pero su uso era muy complejo y conseguir ajustar correctamente una fórmula era una tarea larga y penosa. Con el fin de reducir el tiempo y el coste de producción, la AMS encargó al prestigioso programador Donald E. Knuth que les buscara una solución. El resultado fue un conjunto de utilidades de las cuales la más importante es T_EX; junto con él, se proporcionaban programas para la creación de tipos paramétricos y otras herramientas auxiliares, cuyo número ha aumentado considerablemente con el tiempo.

En esencia, T_EX es un sistema de composición tipográfica de textos de alta calidad y refinamiento, y con funciones avanzadas de automatización. T_EX en sí mismo es un completo lenguaje de programación que permite el marcado lógico de documentos de forma que el formato se pueda generar automáticamente. En Bezos [2003] se expone con más detalle en qué consiste T_EX, sus diferencias con los procesadores de texto y los maquetadores y cómo T_EX se entronca más con los sistemas de fotocomposición, pero la idea fundamental es que mediante órdenes y parámetros de naturaleza tipográfica y la programación se puede hacer casi lo que se desee.⁴

⁴ Ese «casi» es lo que marca la diferencia entre T_EX y los maquetadores: ciertamente T_EX no es apropiado para prensa y diseño gráfico.

Cuadro 1. Tecnologías de edición

<i>Año</i>	<i>Acontecimiento</i>	<i>Relación con T_EX</i>
1980	Knuth desarrolla T _E X y MetaFont.	
1982-1984	Adobe desarrolla PostScript.	El programa dvips permitirá crear archivos PostScript con T _E X.
1984	Apple lanza el Macintosh. Poco después aparece la LaserWriter con PostScript y PageMaker.	
1986	Se aprueba el estándar SGML	Sistemas como Passive T _E X permiten trabajar con él.
1990	Berners-Lee crea HMTL en ordenadores Apple y NeXT.	
1991	Primera versión de Unicode.	La extensión de T _E X llamada Omega permite trabajar con Unicode.
1994	Adobe crea PDF.	Una variante llamada pdftex permite generar directamente archivos PDF.
1998	El consorcio W3 crea XML.	El conjunto de macros <code>xm_ltex</code> permite leer directamente archivos xml.

En el cuadro 1 se dan otros datos que conviene destacar para entender T_EX en su contexto.⁵ Como se ve en él, el mundo de T_EX ha sido lo bastante dinámico como para ir adaptándose a las nuevas tecnologías surgidas en los últimos años, sin perder de vista que esas innovaciones también han tomado ideas de T_EX, como PDF (del formato DVI), XML (con MathML como formato para representar matemáticas), XSL-FO (que ha tenido en cuenta el sistema Omega para documentos multilingües) y Unicode (que tiene presente T_EX en los símbolos matemáticos).

4. Los estilos en T_EX

Una de las diferencias fundamentales de T_EX en la forma de trabajar con el texto, y que comparte con SGML y derivados, es que los cambios no se aplican directamente sobre el documento con el texto, algo que sí hacen programas como Word o PageMaker. Aunque estos últimos programas tienen la capacidad de ser programados (bien con Visual Basic, bien con AppleScript), esto limita seriamente sus capacidades para la automatización.

A cambio, L^AT_EX ha de proporcionar por fuerza una serie de parámetros que han de conformar el aspecto final del documento. Al tratarse de un sistema originario de los Estados Unidos, aunque hoy buena parte de su desarrollo sigue en Europa, no es de extrañar que las normas tipográficas predeterminadas en los estilos estándar tengan una marcada influencia de la tipografía norteamericana. Por tanto, aceptar estos estilos

⁵ Los datos están tomados sobre todo de Goossens y Rahtz [1999].

sin más es inapropiado cuando se trata de componer libros conformes a los criterios hispánicos.

Un paso importante en este sentido lo he dado con una nueva versión del `spanish` (Bezós [2002*b*]) para `babel` (Braams [2001]), que intenta, en la medida de lo posible, adaptar las pautas tipográficas a los usos tradicionales del castellano. Las fuentes consultadas han sido variadas, incluyendo obras clásicas sobre la materia tales como las Morato [1908], Martínez Sicluna [1945], Ramos Martínez [1963] y Polo [1974]. Entre las obras más recientes, las principales son las de Martínez de Sousa [1987, 1992, 1994] y Buen [2000], las únicas modernas que tratan con extensión y seriedad la tipografía. En Bezós [2002*a*] hay un resumen de esas reglas tal y como se aplican en `spanish`.

Entre los cambios más importantes que se introducen están los siguientes:

- Las cifras romanas en minúscula se eliminan, ya que no han tenido tradición en España.
- Las divisiones no sólo se han traducido (*Índice*, *Capítulo*, etc.) sino que se ha tenido especial cuidado en evitar falsos amigos, como *tabla* por «table», ya que en español se ha usado *cuadro*.
- Hay utilidades para que las comillas tengan las formas correctas (« ‘ ’ »), y se añadan automáticamente comillas de seguir, como en este ejemplo:
 - «Se llaman “comillas de seguir” a las que son de cierre, pero se colocan al comienzo de un párrafo.
 - »Este tipo de comillas se emplea cuando se transcribe un texto entrecomillado con más de un párrafo.
 - »En su interior, como de costumbre, se usan inglesas.»
- También hay utilidades para introducir correctamente la coma decimal en fórmulas, así como las funciones `lím` (con acento) o `sen` (en lugar de `sin`).
- Se corrige la norma inglesa de añadir cuadratines tras punto.
- Se añade la ñe en las letras para las enumeraciones.

5. Tres casos

Es imposible describir en un espacio tan breve todas las posibilidades tipográficas de T_EX, pero podemos centrarnos en tres ejemplos que muestran en mayor o menos medida los tres puntos referidos con anterioridad: el primer ejemplo es sobre reglas tipográficas mientras que los dos siguientes ilustran el formato según el formato. Todo ello es posible gracias a que nos encontramos ante un entorno programable.

5.1. Líneas cortas

Un ejemplo de aplicación de reglas tipográficas es la norma que establece que la última líneas del párrafo no ha de ser más corta que la sangría de la primera línea con objeto de evitar un escalón visual. En España también es habitual otra regla por la que se debe evitar que el blanco que quede sea menor que esa sangría, aunque esta última no tiene los efectos tan serios de la anterior y a menudo no se sigue, sobre todo porque en los programas de maquetación hay que hacerlo a mano. T_EX permite automatizar la

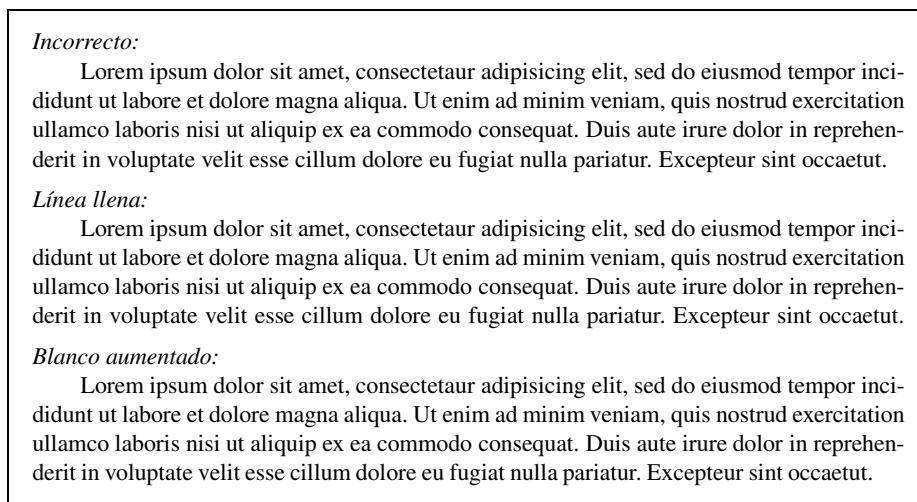


Figura 2. Ajuste de la última línea

aplicación de ambas reglas, y aquí veremos cómo hacerlo con la segunda gracias a una pequeña utilidad escrita por Donald Arseneau.

En corrección tipográfica manual, la norma ha sido llenar la línea, aunque aquí \TeX prueba tanto a llenar como a acortar, y de estas dos soluciones elige la óptima. En el ejemplo que sigue se muestra un párrafo incorrecto y las dos posibles soluciones. De ellas, \TeX opta por la línea llena.

5.2. Poemas

En la composición de poemas se sigue una regla que, de nuevo, no tiene realización directa en los programas de maquetación. En este caso, ni siquiera tendremos la necesidad de recurrir a ninguna extensión a \LaTeX , ya que su entorno `tabular` lo puede tratar directamente, al menos en poemas cortos. La regla que rige es que los versos estarán normalmente alineados por la izquierda, mientras que los poemas estarán centrados con relación al verso más largo. La figura 3 muestra la aplicación de esta regla a un par de poemas, con un posible marcado XML.

Para estos ejemplos, se ha considerado que han de ir sangrados el primer verso de cada poema y los que siguen a versos que terminan con punto, interrogación o exclamación. Las reglas pueden refinarse de ser necesario (por ejemplo, que además el verso empiece con mayúscula), aunque estas tareas pueden dejarse a XSLT o DSSSL y no son específicas de \TeX .

Otro ejemplo relacionado es el teatro en verso, donde hay versos que pueden corresponder a varios interlocutores, tal y como se ve en el ejemplo 4.

```

                ¡Ay, amor,
                perjuro, falso, traidor!
                Enemigo
                de todo lo que no es mal;

                Tú das la caña hermosa,
                de do la miel se acendra,
                por quien desdeña el mundo de los panales;
                tú en urnas de coral cuajas la almendra

<poema>
  <verso>¡Ay, amor,</verso>
  <verso>perjuro, falso, traidor!</verso>
  <verso>Enemigo</verso>
  <verso>todo lo que no es mal;</verso>
</poema>

<poema>
  <verso>Tú das la caña hermosa,</verso>
  <verso>de do la miel se acendra,</verso>
  <verso>por quien desdeña el mundo de los panales;</verso>
  <verso>tú en urnas de coral cuajas la almendra.</verso>
</poema>

```

Figura 3. Composición de poemas

PEDRO CRESPO	¿No hay remedio?
DON ÁLVARO	El de callar es el mejor para vos.
PEDRO CRESPO	¿No otro?
DON ÁLVARO	No.
PEDRO CRESPO	Pues ¡juro a Dios que me lo habéis de pagar!

Figura 4. Composición de teatro en verso

5.3. Títulos

Para ilustrar esto sí usaremos una extensión, denominada `titlesec` y de la que soy autor. Con ella es posible tratar el texto para ver en qué consiste antes de tomar una decisión sobre su formato exacto. En los dos primeros se puede ver cómo se ha medido el título para determinar la arracada (el hueco que se deja en el párrafo). En los dos que siguen primero compone el texto del título, luego lo mide y finalmente añade una línea de topos cuya medida es el ancho del título incrementado en una pica.

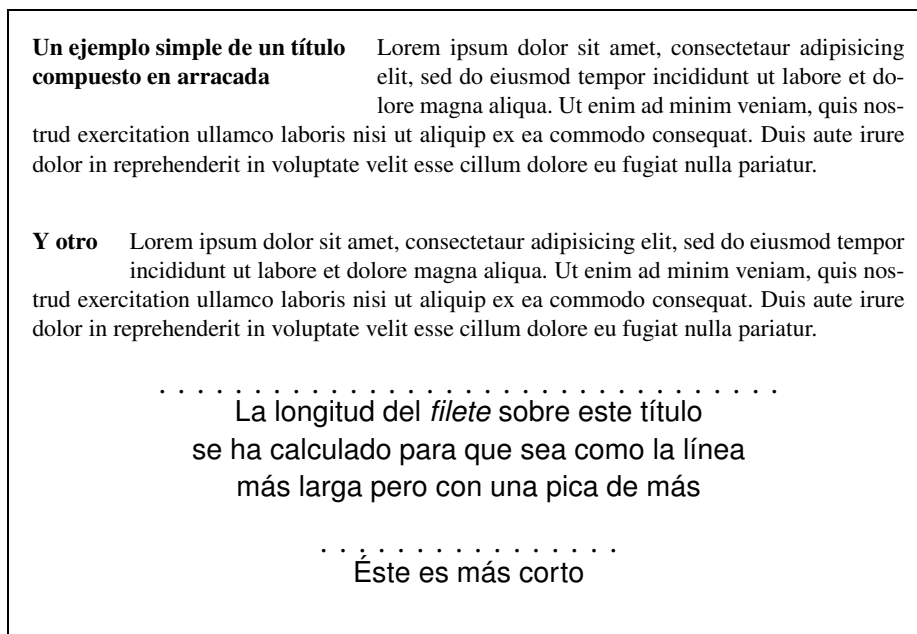


Figura 5. Títulos

Conclusión

El aspecto más importante de la discusión precedente es el uso de $\text{T}_{\text{E}}\text{X}$, más allá de una aplicación puramente mecánica, como un sistema subyacente de composición que permite automatizar una correcta tipografía sin intervención manual. Aquí se ha implementado con $\text{x}_{\text{L}}\text{T}_{\text{E}}\text{X}$, pero de igual forma se podría haber transformado previamente con $\text{P}_{\text{a}}\text{s}\text{s}\text{i}\text{v}\text{e}\text{T}_{\text{E}}\text{X}$, $\text{J}\text{a}\text{d}\text{e}\text{T}_{\text{E}}\text{X}$, $\text{d}\text{b}\text{l}\text{a}\text{t}\text{e}\text{x}$ o Unicorn. Lo realmente relevante no es el aspecto técnico *per se*, sino cómo éste puede ser una herramienta para el aspecto artístico. De lo que se trata, en definitiva, es de recuperar el buen oficio tipográfico, automatizarlo y crear, en palabras de Knuth [1986, p. 303], «masterpieces of the publishing art».

Bibliografía

- Javier Bezos, «Tipografía española con $\text{T}_{\text{E}}\text{X}$ », *T_{\text{E}}\text{X}emplares*, n.º 3, 2002a.
- Javier Bezos, «Imprenta personal con $\text{T}_{\text{E}}\text{X}$ », *PC Actual*, n.º 153, junio 2003, págs. 171-172.
- Javier Bezos, *Estilo spanish para el sistema babel*, 2002b, <http://perso.wanadoo.es/jbezos/archive/spanish.dtx>.
- Johannes Braams, *Babel, a multilingual package for use with $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$'s standard document classes*, 2001, <ftp://ftp.dante.de/tex-archive/macros/latex/required/babel/babel.dvi>, 05/07/2003.
- The Chicago Manual of Style*, Chicago, University of Chicago Press, 14.^a ed., 1993, esp. págs. 333-335.
- Jorge de Buen, *Manual de diseño editorial*, México, Santillana, 2000.

- José Fernández Castillo, *Normas para correctores y compositores tipógrafos*, Madrid, Espasa-Calpe, 1959.
- Michel Gossens, Frank Mittelbach y Alexander Samarin, *The L^AT_EX Companion*, Reading, Addison Wesley, ed. rev., 1994.
- Michel Gossens y Sebastian Rahtz, *The L^AT_EX Web Companion*, Reading, Addison Wesley, 1999.
- Donald E. Knuth, *The T_EXbook*, Reading, Addison Wesley/AMS, 1986.
- Vicente Martínez Sicluna, *Teoría y práctica de la tipografía*, Barcelona, Gustavo Gili, 1945.
- José Martínez de Sousa, *Diccionario de ortografía técnica*, Madrid, Germán Sánchez Ruipérez/Pirámide, 1987. (Biblioteca del libro.)
- *Diccionario de tipografía y del libro*, Madrid, Paraninfo, 3.^a ed., 1992.
- *Manual de edición y autoedición*, Madrid, Pirámide, 1994.
- «Las nuevas tecnologías en el tratamiento de los textos (Los neotipógrafos)», 1998, <http://www.uv.es/barrueco/reb/esp/vol1no1/vol1no1d.html>, 11/06/2003.
- Juan José Morato, *Guía práctica del compositor tipográfico*, Madrid, Hernando, 2.^a ed., 1908 (1.^a ed., 1900, 3.^a ed., 1933).
- Marion Neubauer, «Feinheiten bei wissenschaftlichen Publikationen», *Die T_EXnische Komödie*, parte I, vol. 8, n.º 4, 1996, págs. 23-40; parte II, vol. 9, n.º 1, 1997, págs. 25-44.
- José Polo, *Ortografía y ciencia del lenguaje*, Madrid, Paraninfo, 1974.
- R. Ramos Martínez, *Corrección de pruebas tipográficas*, Mexico, UTEHA, 1963.
- A. Syropoulos y otros, *Digital Typography Using T_EX*, Nueva York, Spinger, 2003.
- «Webster's Style Manual», *Webster's New Encyclopedic Dictionary*, Colonia, Könemann, 1994, págs. 1323-1395.