

El proceso de edición digital en *Artelope* y *CTCE*

Carlos Muñoz Pons
Universitat de València
carlos.munoz-pons@uv.es

Introducción

Una de las cuestiones más importantes a la hora de trabajar en proyectos de investigación es la generación de información que pueda ser compartida entre varios grupos de trabajo. Para ello, es necesario seguir una serie de estándares y normas para que la información pueda ser entendida y tratada por todas las partes fácilmente.

La segunda fase del proyecto *Artelope*²⁶⁰, llamada *Artelope II*, y la Colección de Teatro Clásico Europeo (CTCE) tienen como objetivo principal la generación de documentos en formato electrónico de los textos de las obras que incluyen. En el caso de *Artelope*, de las obras de Lope de Vega, y en el caso de la Colección de Teatro Clásico Europeo, de títulos de varios autores europeos en distintos idiomas. Al compartir objetivo, ambos proyectos comparten también los métodos y las herramientas para realizar esta edición digital.

Para la codificación de estos textos hemos seguido el estándar TEI-XML. Esta codificación nos permite clasificar cada uno de los elementos que intervienen en las obras de teatro. Con las herramientas adecuadas, podremos realizar acciones y obtener características del texto que no nos permitirían otros formatos electrónicos.

Como cualquier fichero XML, los ficheros TEI-XML deben estar *bien formados*, es decir, seguir las reglas básicas que cumplen todos los ficheros XML, y ser *válidos*, seguir las normas dictadas por su DTD (definición de tipo de documento) o por su esquema. Como esto puede ser un trabajo complicado y tedioso para investigadores poco familiarizados con este tipo de ficheros, se han creado herramientas que permiten la generación correcta de estos ficheros de forma automática.

²⁶⁰ Oleza, Joan et al. *Base de Datos y Argumentos del teatro de Lope de Vega. ARTELOPE*.

En este artículo vamos a describir cómo es el proceso de edición digital en ambos proyectos, desde que tenemos un texto en un formato *clásico* hasta que obtenemos un fichero en formato digital.

El artículo constará de 3 apartados principales:

- Una pequeña introducción mostrando el proceso de edición de forma general.
- Una descripción más detallada de cada una de las fases del proceso.
- Un apartado donde explicaremos el trabajo futuro de los proyectos.

Descripción general del proceso de edición



Figura 1. Descripción general del proceso de edición

En la figura superior vemos de forma genérica cómo es el proceso de edición. En primer lugar, partimos de un fichero en formato clásico. Estos archivos pueden ser una obra en papel o un fichero en formato digital (por ejemplo: un archivo PDF o archivo Word). Sobre este documento original, realizamos el proceso de edición digital propiamente dicho y obtenemos un fichero en formato electrónico.

Este fichero que obtenemos podemos consultarlo desde un ordenador o desde una tableta electrónica, como veremos más adelante. Por supuesto, esta edición digital *final* incluye, además del propio texto, información y funcionalidades adicionales a la obra.

El proceso de edición digital que vemos en la parte central de la figura se puede desglosar en las fases que vemos a continuación:

- Digitalización del texto.
- Marcación previa del texto.
- Incorporación a la base de datos.
- Edición desde la base de datos.
- Exportación a TEI-XML.
- Transformación a otros formatos.

Descripción detallada de las fases del proceso de edición

En este punto vamos a describir cada una de estas fases de forma más detallada.

Digitalización del texto



Figura 2. Digitalización del texto

El primer paso es la digitalización del texto. El objetivo es obtener un fichero de texto sin formato (texto plano) con el contenido de la obra. Como hemos comentado anteriormente, podemos partir de una obra en papel o de una obra en formato electrónico. Estos documentos pueden ser proporcionados desde fuentes, como la Biblioteca Virtual Miguel de Cervantes con quienes tenemos un convenio.

La razón de generar un fichero de texto sin formato es garantizar que la información sea lo más homogénea posible y evitar problemas de visualización y clasificación de los elementos en las siguientes fases del proceso.

Marcación previa del texto

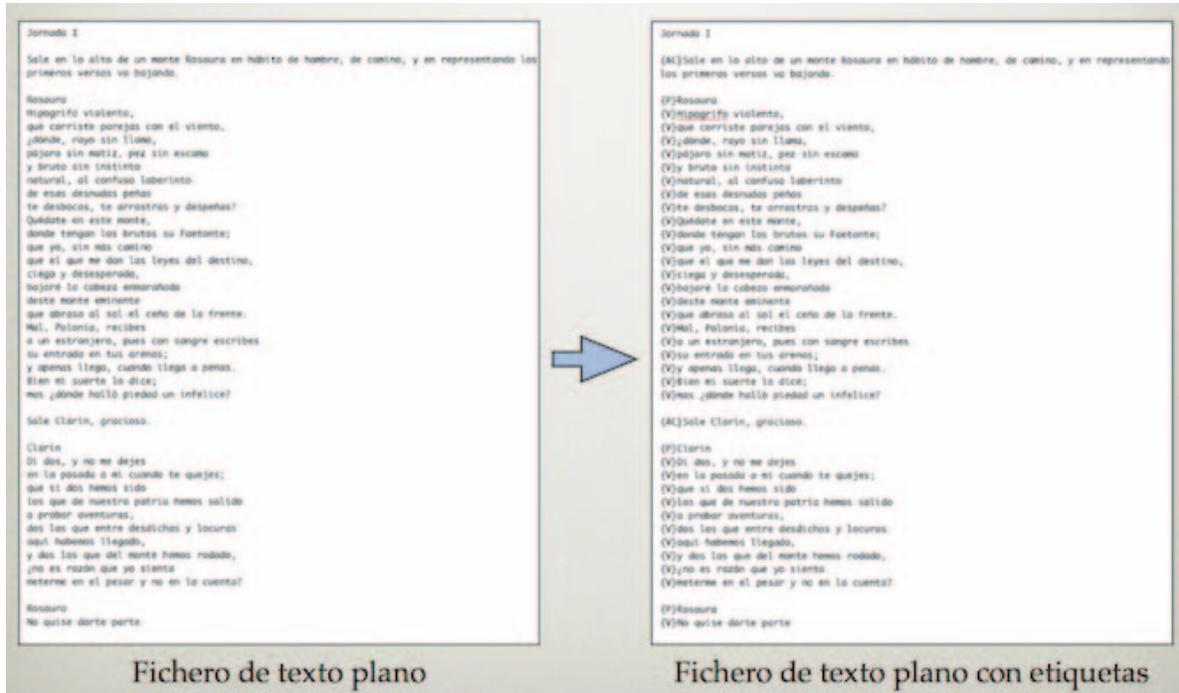


Figura 3. Marcación previa del texto

Una vez tenemos el contenido de la obra en un fichero de texto sin formato, realizamos sobre él una marcación previa a la incorporación a la base de datos. Esta marcación consiste en la incorporación de una serie de etiquetas sobre el texto para clasificar los elementos de la obra: acotaciones, nombre de los personajes, versos, etc. Esto facilita la inserción de contenido en la base de datos y la clasificación de los distintos elementos del texto en la misma.

```
{AC}Sale en lo alto de un monte Rosaura en hábito de hombre, de camino, y en representando los primeros versos va bajando.

{P}Rosaura
{V}Hipogrifo violento,
{V}que corriste parejas con el viento,
{V}¿dónde, rayo sin llama,
```

Figura 4. Ejemplo de etiquetado

Como vemos en el ejemplo de la imagen anterior, hemos añadido delante de la acotación la etiqueta {AC}, delante del nombre del personaje, la etiqueta {P}, y delante de los versos, la etiqueta {V}. Más adelante veremos cómo se clasifican cada uno de los elementos en sus campos correspondientes.

Incorporación a la base de datos

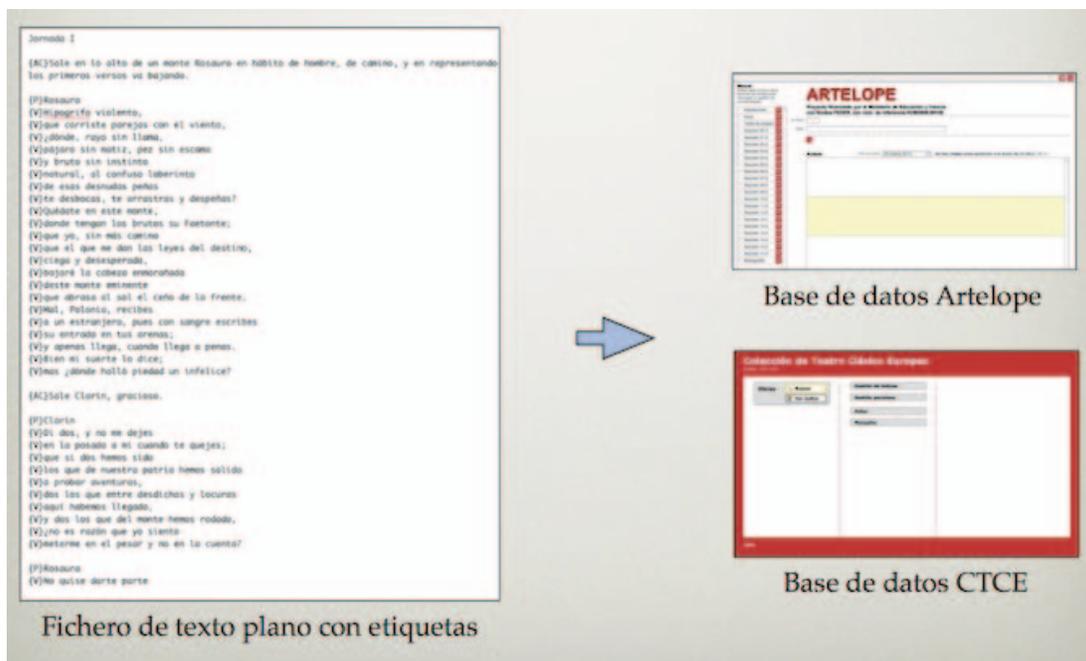


Figura 5. Incorporación a la base de datos

Una vez tenemos el texto con esta marcación previa ya podemos incorporarlo a una de las bases de datos FileMaker. Hay que destacar que ambas bases de datos, Artelope y CTCE, utilizan el mismo sistema de edición. Esto conlleva 2 ventajas principales:

- Un mismo texto puede incorporarse en una base de datos o en otra, sin variar la marcación utilizada. Incluso podemos editar una obra en una base de datos y trasladarla, una vez finalizada la edición, a la otra fácilmente.
- Un investigador que ya esté trabajando en una base de datos puede trabajar en la otra sin tener que aprender de nuevo cómo funciona todo el sistema de edición.

Actualmente, ambas bases de datos se encuentran alojadas en un servidor en el Centro de Cálculo de la Universitat de València. Esto nos aporta algunas ventajas interesantes:

- Acceso ininterrumpido las 24 horas del día.
- Excelente rendimiento de red, que permite que investigadores de otros países accedan a las bases de datos con una velocidad adecuada.
- Un nivel de seguridad extra, mediante el uso de cortafuegos y otros sistemas de seguridad.

Edición desde la base de datos



Figura 6. Edición desde la base de datos

Una vez el texto de la obra está introducido en la base de datos, el investigador trabaja desde una interfaz como la de la imagen anterior. En este caso, la captura está sacada de la base de datos CTCE, al igual que las siguientes que mostraremos, aunque en la base de datos Artelope se utiliza una interfaz muy parecida.

Descripción de la interfaz



Figura 7. Zonas de trabajo y de visualización

En la parte superior (zona azul) se encuentra la zona de trabajo, donde el investigador realiza la mayoría de las acciones. En la parte inferior (zona amarilla) se encuentra la zona de visualización, donde el investigador puede ir viendo el resultado de la edición que está realizando en formato HTML o TEI-XML.

Cada una de las filas de la zona de trabajo representa una línea del texto que estamos editando. Si nos fijamos en el ejemplo de la figura anterior, podemos observar como la primera fila se corresponde con la acotación, la segunda contiene el nombre de personaje y el primer verso, etc.

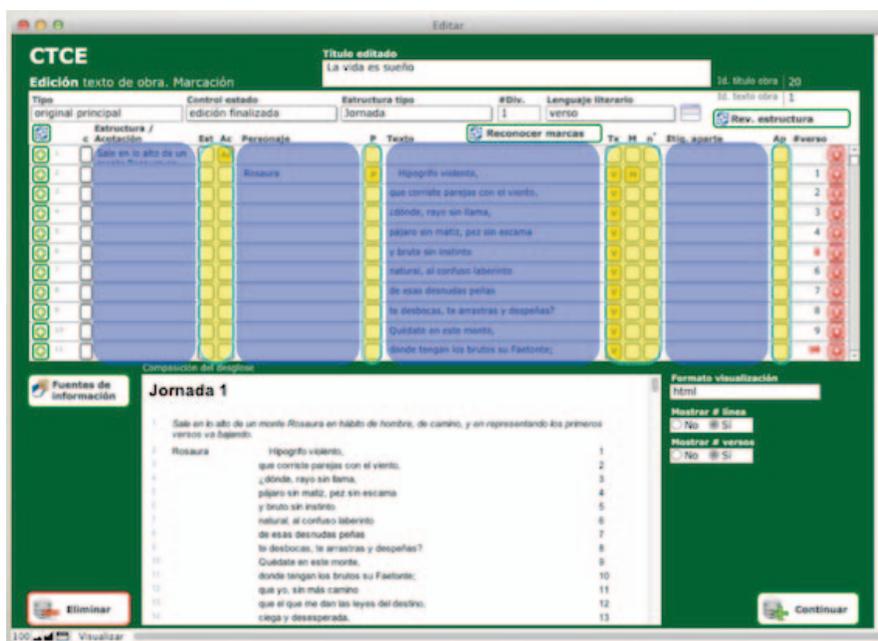


Figura 8. Campos de texto y botones

Todas las líneas están formadas por campos de texto (zona azul) y por botones (zona amarilla). En cada uno de los campos de texto se inserta la información. Dependiendo del tipo que sea, irá en una columna u otra: el texto de la acotación en la primera columna, el nombre de personaje en la segunda, el texto del verso en la tercera y la etiqueta de los apartes en la cuarta. La información se inserta en cada columna, dependiendo de la etiqueta que hayamos utilizado en el apartado de 'Marcación previa del texto'. Con los botones podemos realizar acciones adicionales sobre estos campos de texto.



Figura 9. Añadir y eliminar líneas

Es posible que al texto utilizado para incorporar en la base de datos le falte alguna línea. En cualquier momento podemos añadirla pulsando el botón de la parte izquierda la fila de la interfaz. Igualmente, podemos eliminar una línea que no necesitemos pulsando el botón de la parte derecha.

Operaciones adicionales

Como hemos comentado hace un momento, pulsando en los botones centrales de cada línea podemos realizar acciones sobre los campos de texto y completar o añadir más información. Vamos a ver algunos ejemplos de estas acciones:

- **Tipo de acotación**

Podemos definir por ejemplo el tipo de acotación. Para ello, pulsamos sobre el botón Ac (círculo rojo) que vemos en la figura anterior. Cuando lo hagamos, nos aparecerá una nueva ventana. En esta ventana encontraremos un desplegable que nos permitirá seleccionar si la acotación describe una entrada, una salida, una descripción del lugar, etc.



Figura 10. Tipo de acotación

- **Tipo de línea**

También podemos definir el tipo de verso. Para ello, pulsamos esta vez sobre el botón V (círculo rojo). Se nos abrirá una nueva ventana en la que podremos definir el tipo de línea que estamos editando: si es un verso, si es un verso truncado, si es prosa, etc.

A la hora de incorporar el texto, la línea se habrá clasificado ya como prosa, verso, etc. Pero en algunos casos es necesario un ajuste más fino del tipo de línea.

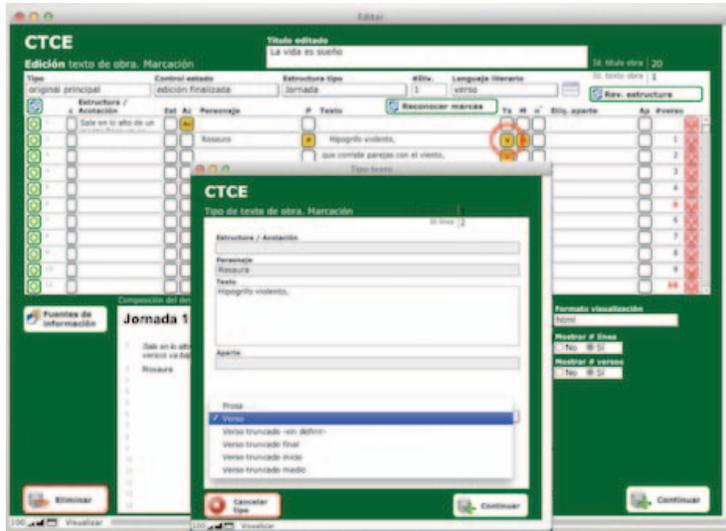


Figura 11. Tipo de línea

- **Tipo de estrofa**

Podemos definir también el tipo de estrofa. Para ello, pulsamos sobre el botón M (círculo rojo) y nos aparecerá una nueva ventana con el listado de posibles métricas. Una vez seleccionemos una, estaremos indicando que en ese verso empieza un determinado tipo de estrofa. Además, al marcar el tipo de métrica, se añade una pequeña sangría al texto del verso.



Figura 12. Tipo de estrofa

- **Tipos de visualización**

Como hemos comentado, todas estas acciones que hemos ido realizando lo que provocan es que el texto se vaya modificando en la parte inferior de la pantalla (zona de visualización). De esta forma, podemos ir viendo cómo va quedando el texto mientras editamos.

Estas modificaciones se realizan en tiempo real, por lo que cualquier cambio que realicemos en la parte superior, tendrá repercusión instantáneamente en la visualización inferior.

En las capturas que hemos visto hasta este momento, hemos utilizado una visualización en el formato más clásico, la visualización web. Como el objetivo del proyecto es la generación de ficheros TEI-XML de los textos de las obras, tenemos la posibilidad de ver cómo va quedando el texto también en ese formato.



Figura 13. Visualización en HTML

Para ello, debemos seleccionar en el desplegable que vemos en la imagen (círculo rojo) la opción de TEI-XML y automáticamente veremos el texto de este formato.

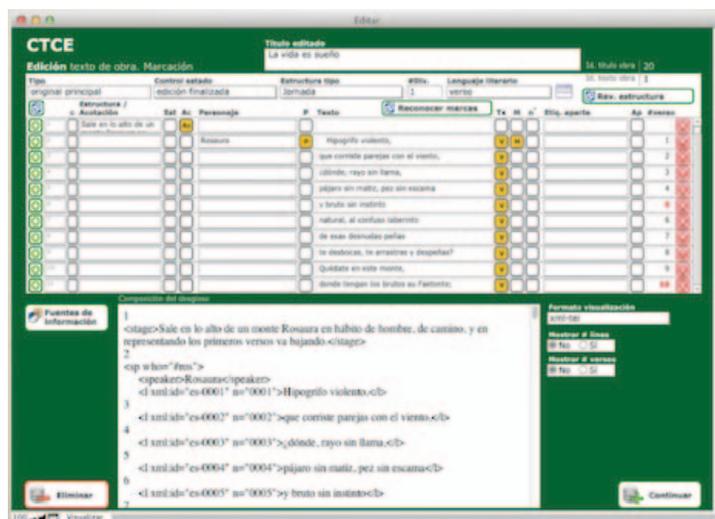


Figura 14. Visualización en TEI-XML

Generación del fichero TEI-XML

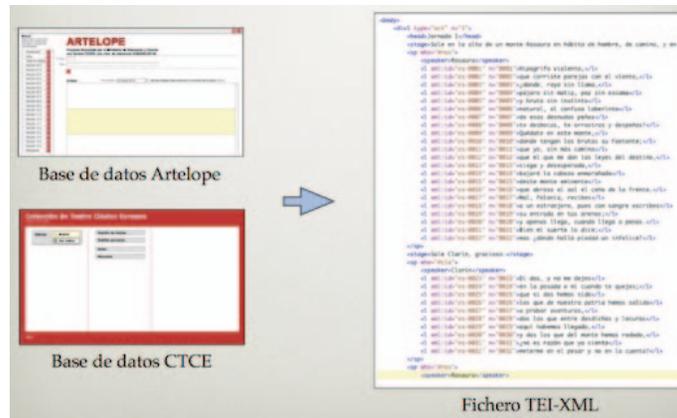


Figura 15. Generación del fichero TEI-XML

Una vez hemos terminado de editar el texto desde la base de datos, el siguiente paso es exportar el fichero TEI-XML. Este fichero ya contiene todo lo necesario para trabajar con él: contadores de líneas y versos, estructuras de estrofas, códigos de personajes, etc. Hay que destacar que la generación del documento TEI-XML se realiza de forma automática y en ningún momento el investigador trabaja ni edita directamente el documento XML.

Esto tiene la ventaja de que el fichero TEI-XML tiene un formato homogéneo para todos los textos, por lo que garantiza que la conversión a otros formatos y la utilización de las herramientas no suponga ningún problema.

Transformación a otros formatos

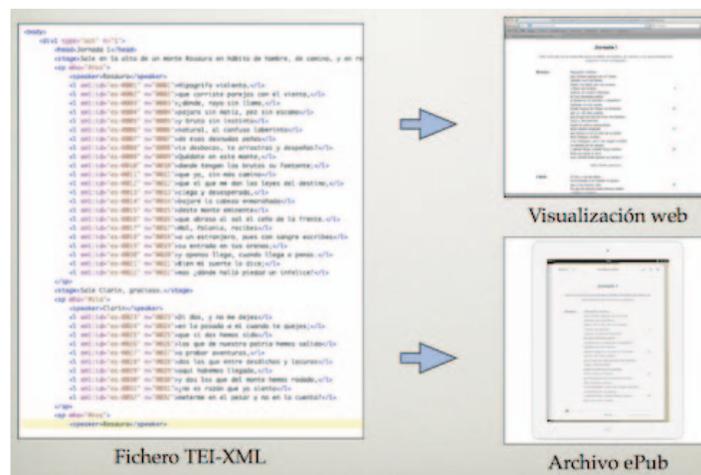


Figura 16. Transformación a otros formatos

Una vez exportamos el fichero TEI-XML, tenemos que transformarlo en un formato más visual y sencillo. Utilizar el formato TEI-XML en los ficheros de los textos tiene muchas ventajas, pero éstas están centradas sobre todo cuando son máquinas las que procesan la información. Sin embargo, su lectura por un investigador puede resultar complicada.

Es por esto que nuestros proyectos realizan 2 tipos de transformaciones en este punto:

- Publicación en formato web, pensado para su uso con un ordenador.
- Publicación en formato ePub, pensado para su uso con una tableta o lector de libros electrónico.

Ambos métodos de transformación los hemos desarrollado nosotros mismos. En la siguiente tabla vemos una serie de características propias de cada formato:

Formato web	Formato ePub
<ul style="list-style-type: none"> • Imprimir la obra. • Realizar búsquedas utilizando el propio navegador. • Copiar y pegar fragmentos del texto. • Navegar de forma rápida a lo largo del texto. 	<ul style="list-style-type: none"> • Índices automáticos. • Búsqueda en los diccionarios integrados en el dispositivo. • Búsqueda de ocurrencias y contextos. • Posibilidad de resaltar palabras y frases.

En ambos casos ofrecemos el texto de la obra junto con las estadísticas de la misma. Algunas de estas estadísticas son:

- Número total de versos.
- Distribución de versos por actos.
- Distribución de versos por personajes.
- Distribución de versos por tipo de estrofa.
- Número de versos en aparte.

Trabajo futuro

En este punto vamos a describir brevemente cuales son los objetivos futuros de ambos proyectos, en lo que se refiere a la edición de textos.

Portal para publicar los textos editados

Por un lado, estamos desarrollando un portal donde iremos publicando los textos que hemos trabajado. En estos momentos ya podemos generar las obras en los formatos comentados anteriormente, pero no podemos ofrecerlas fácilmente al público.

Es necesario entonces crear una plataforma de distribución. En ella se ofrecerán de forma periódica las obras que vayamos terminando de editar. Además, este portal tendrá conexión con la web de [Artelope](#) de forma que desde la propia ficha de la obra podamos acceder directamente a los textos en ambos formatos.

Herramientas de análisis y tratamiento avanzado de los textos de las obras

Por otro lado, estamos creando herramientas de análisis y tratamiento avanzado de los textos de las obras. En la versión web queremos obtener estadísticas más avanzadas, comparación de estadísticas de distintas obras, ofrecer la posibilidad de obtener las intervenciones exclusivas de los personajes, una visualización avanzada de las notas del texto, etc.

Además, estudiamos la posibilidad de crear una aplicación externa, para ordenador y/o tableta, con la que trabajar directamente con los textos en formato TEI-XML.

Edición de textos en paralelo (CTCE)

Y por último, tenemos la fase más ambiciosa del proyecto de la Colección de Teatro Clásico Europeo: la edición de textos en paralelo con las distintas traducciones de una misma obra. Una vez realizado el proyecto, ofreceremos las 4 *versiones* de la obra, una por traducción, enlazadas unas con otras, de forma que podamos ver las correspondencias de los texto entre las distintas versiones.