

Archivo a procesar

***Archivo:**

Formato de entrada

Etiquetación de entrada

Pasos a llevar a cabo

Pasos

* denotes required field

GRUPO DE INVESTIGACIÓN VUM

Utilidad de Transformación de Corpus a Documentos XML (UTITRA)

UTITRA ofrece la transformación en documentos XML válidos de documentos transliterados siguiendo procedimientos inspirados en SGML y bastante extendidos entre los editores de corpus orales del español. El proceso sigue nueve estadios o pasos de depuración y adaptación. El documento meta tiene dos versiones:

- en el paso 9 aparece un documento de presentación en cajas con, en su caso, indicación de errores.
- en el paso 8 se muestra el documento XML.

Recomendaciones

1. Se recomienda acceder a la dirección y trabajar en el entorno del navegador libre Mozilla Firefox. En los demás navegadores el sistema puede no funcionar correctamente.
2. La utilidad admite documentos en Microsoft Word y en texto plano.
3. Los documentos fuente deben ser conformes a las dos siguientes propuestas:
 - VUM: [Ávila Muñoz, Antonio, Matilde Vida y M.^a de la Cruz Lasarte \(2007\): Propuesta de transliteración y etiquetado del macrocorpus PRESEEA. Universidad de Málaga \[última consulta: 11/11/2007\]](#)
 - PRESEEA: [Moreno Fernández, Francisco \(2008\): PRESEEA. Marcas y etiquetas mínimas obligatorias. Vers. 1.1. 12-02-2008. Universidad de Alcalá \[última consulta 18-03-09\].](#)
4. El sistema es incapaz de reconocer los turnos si la letra inicial que introduce los turnos de hablante va seguida solo por dos puntos (H: E: I:). Es necesario cambiar este procedimiento de identificación de hablantes por el siguiente: inicial seguida de punto y dos puntos (H.: E.: I:).

Esta modificación deberá realizarse en el documento fuente antes de introducirlo en el sistema.

5. Debe controlarse, en la medida de lo posible, que toda etiqueta abierta tenga cierre. En caso contrario, el etiquetado se expandirá más allá de lo requerido.
6. Las anotaciones (como paréntesis, barras de pausa, signos de interrogación, etc.) que no vayan entre paréntesis angulares serán consideradas como texto por el sistema.
7. En el caso de la localización de errores, debe corregirse el documento fuente y volver a procesar.