

TAligner 3.0

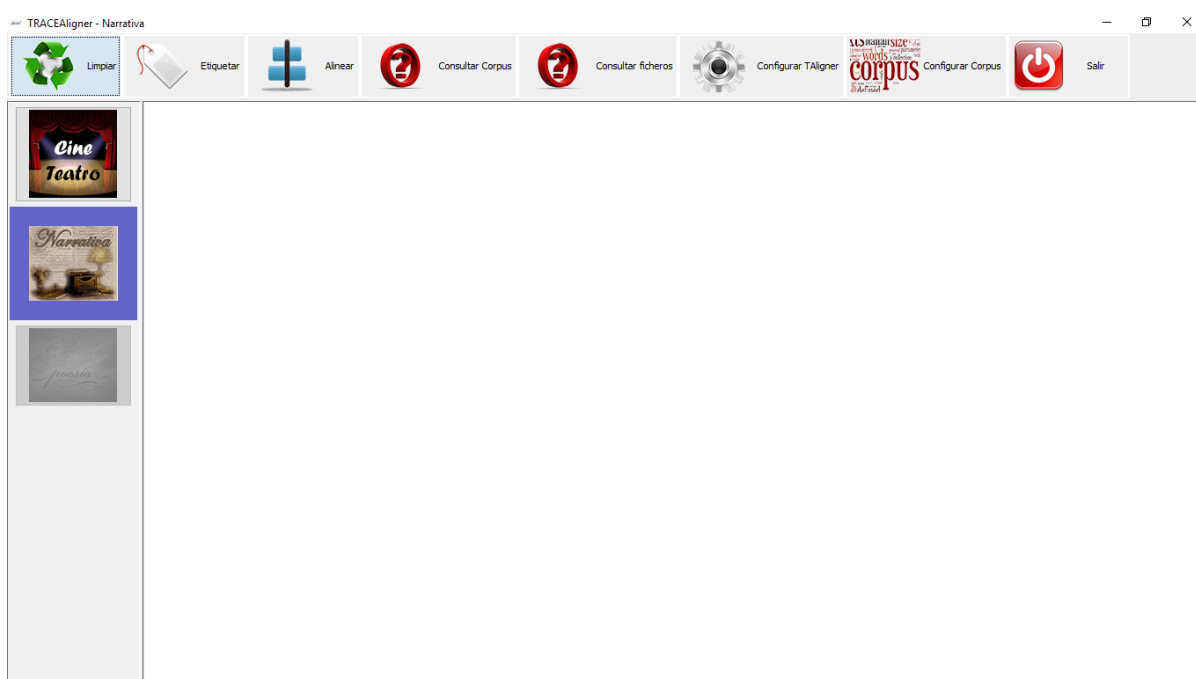
TAligner is a corpus alignment tool developed by the research group TRALIMA/ITZULIK (University of the Basque Country UPV/EHU). This tool, which has been inspired on TRACE Aligner (University of León), is written in Java and equipped with an embedded database that lets the tool browse an aligned parallel corpus.

<https://addi.ehu.es/handle/10810/34436>

<http://corpusnet.unileon.es/herramientas-tecnicas>

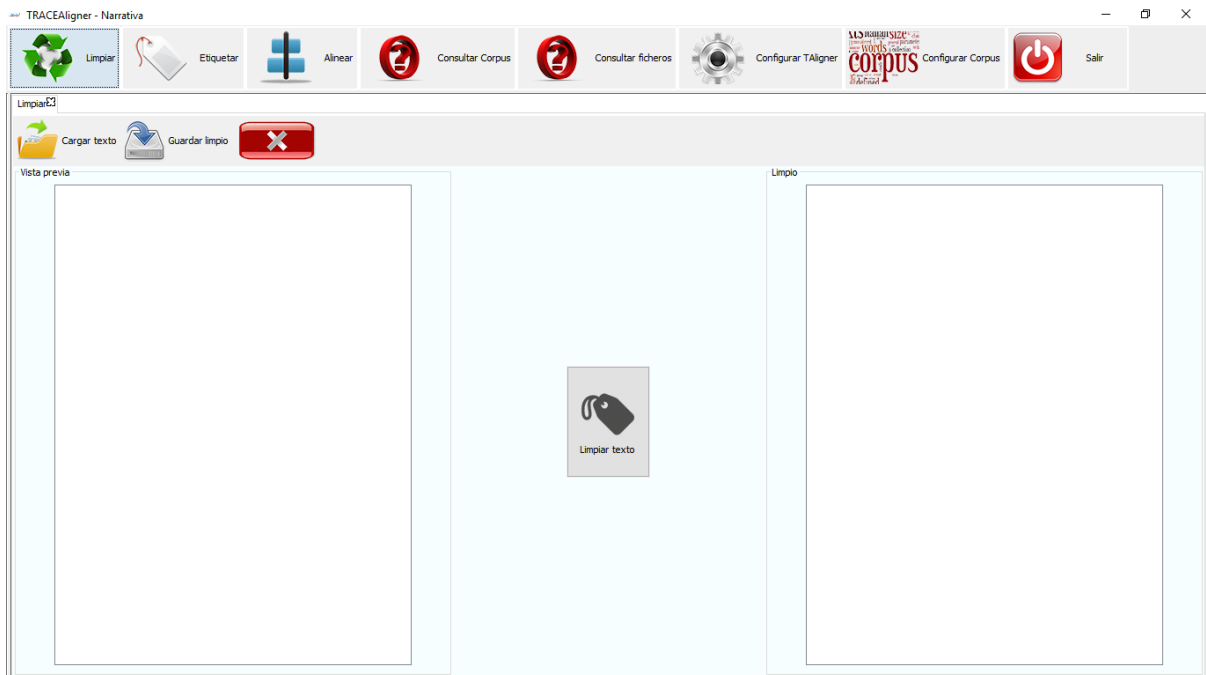
Below follows a short user manual.

1. Text type selection



To start using the tool, we first need to choose the text type we want to work with. This step is essential given that the tagging and alignment procedures differ per text type. On the left side of the screen there is a menu to choose the text type from. The default choice is 'narrativa/narrative texts'. However, once we have selected a specific type, the tool remembers that choice for ensuing working sessions. It is important to keep in mind that the text type menu is displayed only when no other windows are open.

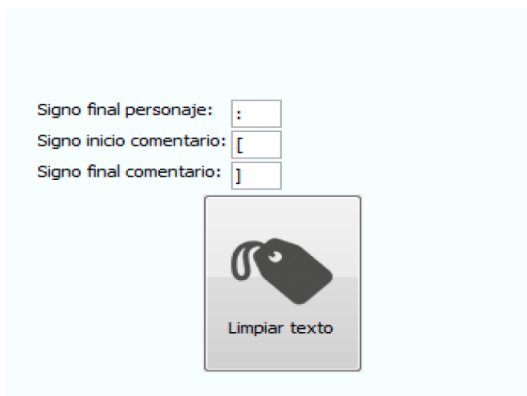
2. Text cleaning



Cleaning our texts is essential prior to using the tool; it has to be done so that the texts take on a standard format that enables their tagging and alignment to be done as fine as possible (few mistakes occur). The procedure is simple:

2.1. First we click on "Cargar texto/Upload text". We will get a pop-up window with a list of texts to choose from. The text selected is then displayed on the left of the screen.

2.2. In the case of theatre texts, we get to see three different text boxes to adjust the cleaning of the texts. Within such text boxes, we see the punctuation marks used to determine where the name of a character finishes, as well as the start and end of stage directions. If the text to work with featured different marks, we should include them within the text boxes mentioned above before cleaning them.



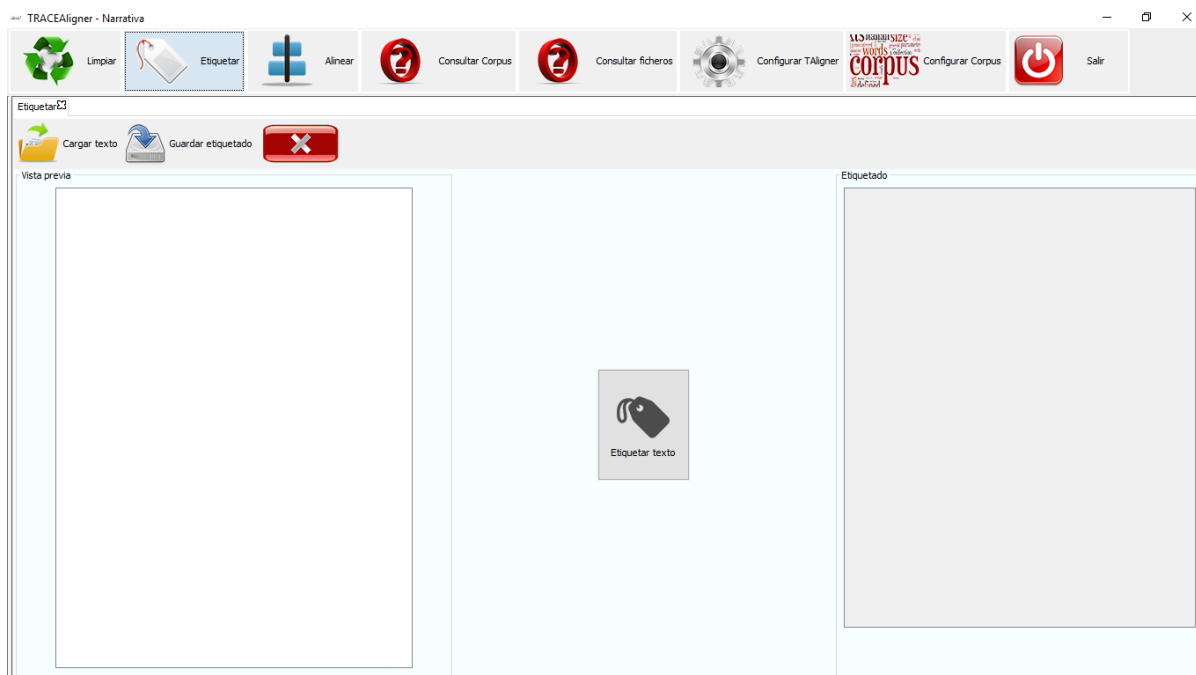
2.3. Once the text is uploaded, we will click on "Limpiar texto/Clean text". The cleaned text is then shown on the right side of the screen.

2.4. Finally, we should save the cleaned text so that it can be tagged later.

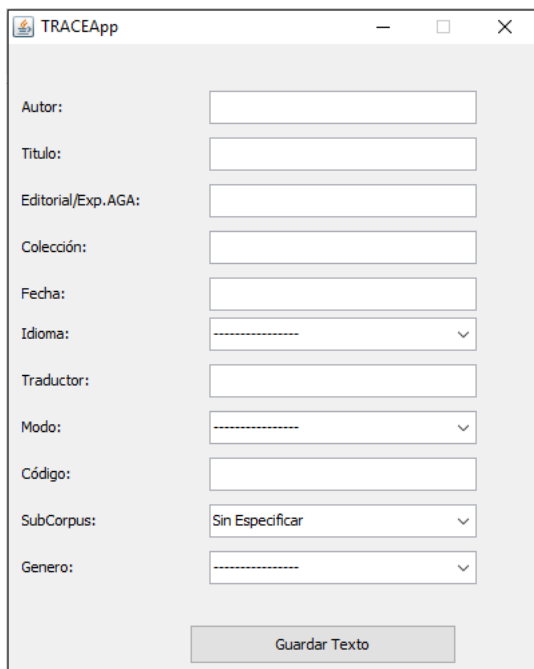
3. Text tagging

Once cleaned, every text may be tagged. This step is mandatory so that it can be aligned with other texts. The procedure resembles that of text cleaning: first we upload the text, then we tag it and finally save it where pertinent.

On the left hand side of the screen we see the text to be tagged. On the right, we will see the text tagged.



When we are about to save a tagged text, we get another pop-up window asking the user to enter the text metadata. Such information is necessary for browsing the texts; therefore, the more metadata, the better.



The image shows a screenshot of a window titled "TRACEApp". The window contains a form for entering text metadata. The fields are as follows:

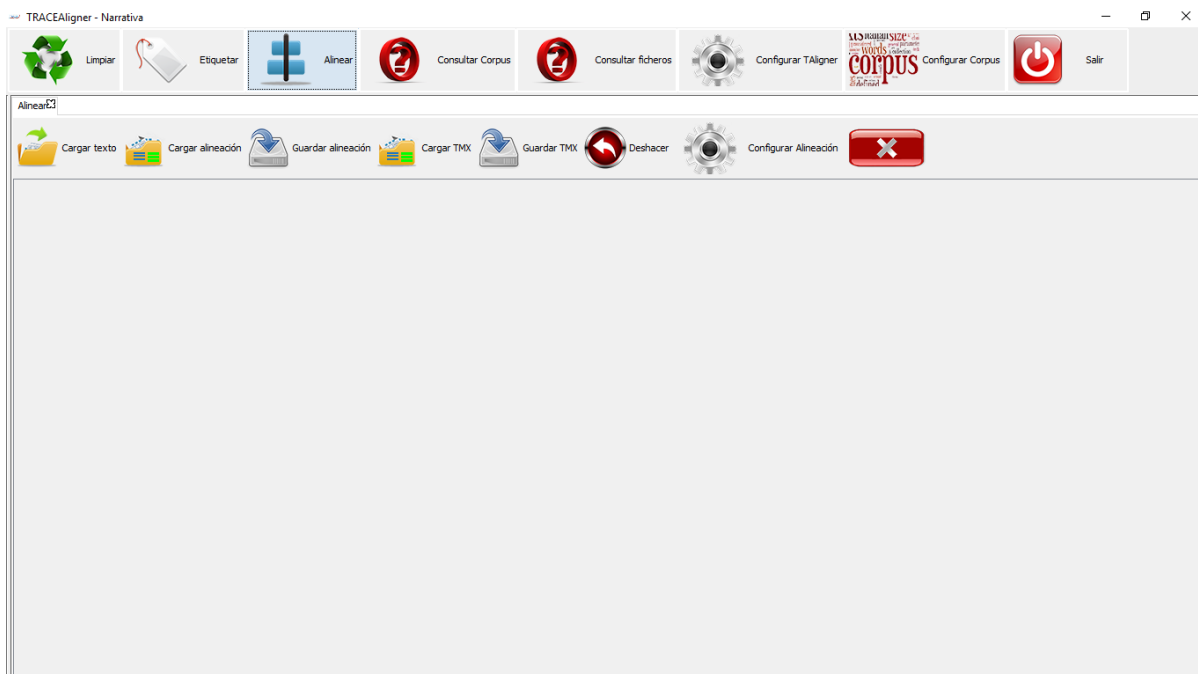
- Autor:
- Título:
- Editorial/Exp.AGA:
- Colección:
- Fecha:
- Idioma:
- Traductor:
- Modo:
- Código:
- SubCorpus:
- Genero:

At the bottom of the form is a button labeled "Guardar Texto".

4. Text alignment

The alignment cannot be done unless we have tagged two -or more- texts. As observed on the screen, the menu is more complex, with the following options:

- **Upload text:** a new text is added to the alignment process.
- **Upload alignment:** a set of two or more aligned texts is uploaded.
- **Save alignment:** the current set of texts under alignment is saved.
- **Upload TMX:** this option lets us upload an alignment as a TMX file. This is a standard format that lacks the texts' metadata. This is why it is recommended to use the TAligner own format by clicking on "Cargar alineación/Upload alignment". Uploading TMX turns out useful in case we want to save our aligned texts using other tools.
- **Save TMX:** This option lets us save the current alignment job as TMX, without metadata being saved. This is recommended only if we want to export the alignment to other tools. However, if we want to save a current alignment job unfinished to resume it later on, it's better to use the "Guardar alineación/Save alignment" option.
- **Undo:** every time we click on this button, the last step is undone.



Together with the aforementioned options, once we see on the screen that all the texts have been uploaded, we can carry out the actual alignment. To this end, we will right click on the texts we want to align. Immediately after this, a new menu is displayed.

Combine: we select as many cells as we want and merge them all into just one by clicking on "Combinar/Combine"

Insert blank: this option adds a blank cell above the cell which we have selected.

Split: this option lets us divide the content within one cell into two parts. In order to do this, a new screen with two cells per chosen cell is displayed. What we must do is writing down the content in the two new cells the way we want it to appear.

Edit: this option lets us change the content within a cell.

Delete: the chosen cell is deleted.

Save XML: this option lets us save as an XML file the text in the column we have clicked on. This way, we save a text individually, together with all the changes made.

Edit metadata: this option lets us change all the metadata in a text.

Add to corpus: this step should be done only after the alignment has been completed. This option lets us add to the corpus the text from the column we have clicked on. This could take long to complete, mainly if the texts are large. It has to be done with each and every aligned text we want to add to the corpus. Even though it is not necessary, it is recommended to add to the corpus first the original text, so that the various translations can be related to that source text.

Delete column: by clicking on this button, we are able to delete a whole text from the alignment screen.

Adjust cell hight: this option makes it possible that all the cells are adjusted to their content.

5. Corpus browsing

This screen lets us browse the texts added to the corpus. On the upper menu, we will specify our search input. Then we click on "Filtrar/Filter", retrieving the hits, which can be exported as an Excel spreadsheet. To do this, we need to click on "Exportar/Export"

TRACEAligner - Narrativa

Limpiar Etiquetar Alinear Consultar Corpus Consultar ficheros Configurar TAligner Configurar Corpus Salir

Consultas corpus

Consultar corpus										FILTRAR
Idioma de búsque...	-----	Modo de búsqueda:	Exacto	Texto a buscar 1:						SALIR EXPORTAR
Modo de búsqueda:	Exacto	Texto a buscar 2:		Distancia a Texto 1:	Cualquiera					
Modo de búsqueda:	Exacto	Texto a buscar 3:		Distancia a Textos...	Cualquiera	Cualquiera				
Autor:	Todos	Traductor:	Todos	Título:	Todos					
Modo:	Cualquiera	Género:	Cualquiera	Buscar traducciones:	<input checked="" type="checkbox"/>	No diferenciar Ma...	<input checked="" type="checkbox"/>			
Cond. Especial:	Todos	Antes y después:	0	SubCorpus:						

Nº Coincidencias: 0

6. File browsing

Below is shown the screen for file browsing.

The screenshot shows the TRACEAligner - Narrativa application window. The title bar reads "TRACEAligner - Narrativa". The menu bar includes: "Limpiar" (Clean), "Etiquetar" (Tag), "Alinear" (Align), "Consultar Corpus" (Consult Corpus), "Consultar ficheros" (Consult files), "Configurar TAligner" (Configure TAligner), "Configurar Corpus" (Configure Corpus), and "Salir" (Exit).

The main interface is titled "Consultas ficheros" (File queries) and features a "Cargar textos" (Load texts) button with a red 'X' icon. Below this, there are two main sections: "Ficheros seleccionados" (Selected files) and "Resultado" (Result).

The "Ficheros seleccionados" section contains a list box labeled "Ficheros" which is currently empty.

The "Resultado" section displays a table with the following structure:

Palabra	Cantidad	Palabra	Cantidad

Below the table, the following statistics are shown:

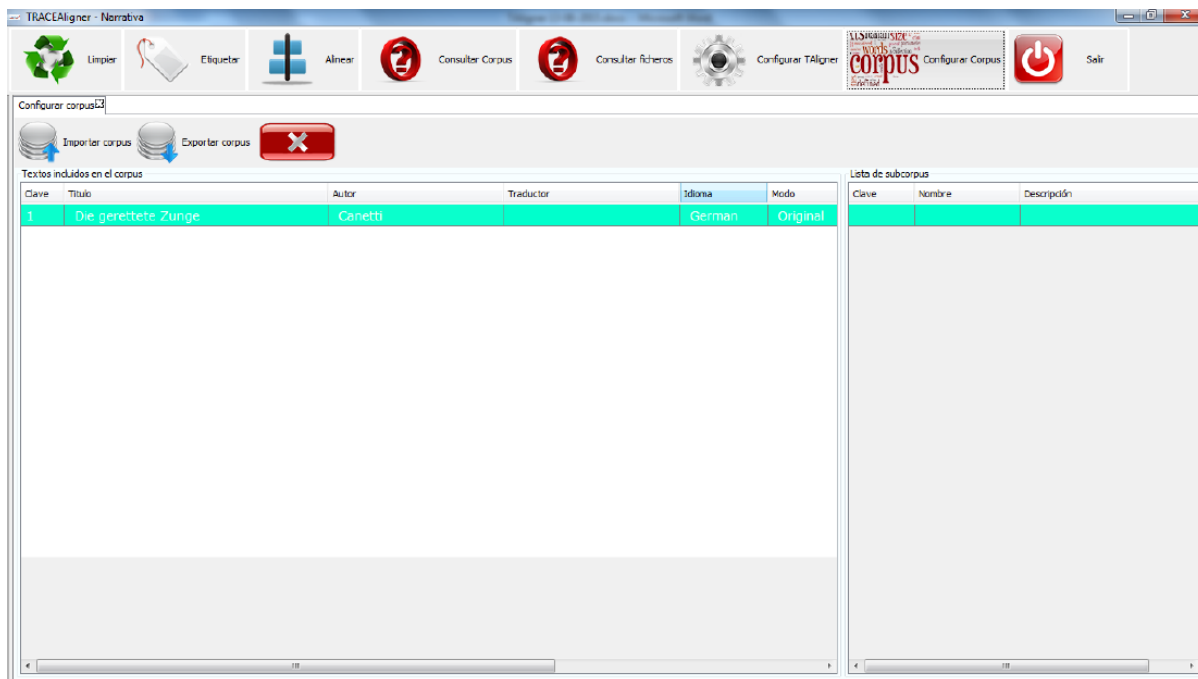
- Total palabras: 0
- Total palabras diferentes: 0
- Media aparición por palabra: 0.0
- Palabras que cumplen filtro: 0
- Media: 0

On the right side of the interface, there are several options and filters:

- Diferenciar may/min:
- Quitar números:
- Mínimo de repeticiones: [input field]
- Máximo de repeticiones: [input field]
- Consultar
- Exportar

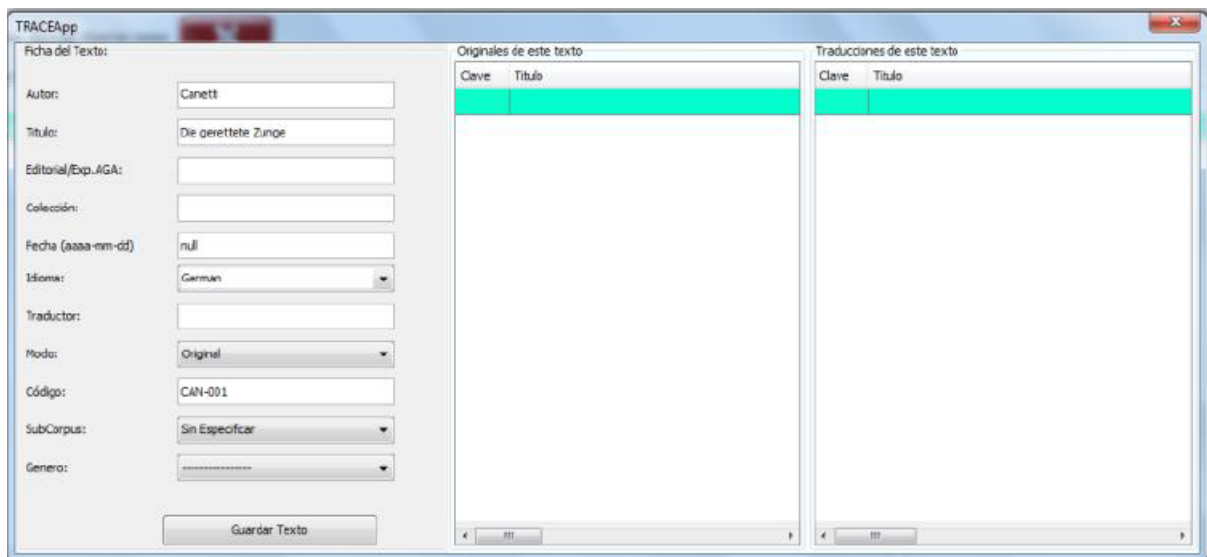
7. Setting up the corpus

The interface for setting up the corpus shows two tables. The one on the right (smaller) displays a list with all the sub-corpora that have been created so far. This is helpful for us to group texts and then 'filter' our searches. To create or delete a subcorpus we need to do right click on the table lines.



The table on the left is essential in that it displays what texts have been added to the corpus. Such texts have been transferred from the alignment window. To remove a text from the corpus, we do right click on the text to delete and then we choose the option "eliminar/delete".

Both by clicking on a given text and by doing right click while choosing "Editar Metadatos and relaciones/Edit metadata and relations", we get to see a new interface:



Here we can change any metadata of a corpus text, without having to remove it first and then add it again. In addition, and most importantly, this interface shows the relations between texts.

We have two tables in the window. One of them is labelled "Originales de este texto/Source texts" and the other one "Traducciones de este texto/Translations". In the first one we can add (or delete) as many texts as we want to label as original versions of the text that is shown on the screen. Likewise, in the second table we may mark what texts are translations of the one shown on the screen.

It is unnecessary to encode the relations twice. In other words, if we have two texts: "Text 1" and "Text 2", and 1 is a translation of 2, we could specify it either in 1 or in 2. Whether we specify this relation in the metadata of 1 or in those of 2, the information is either way shown in both of them.