Quickstart – Importing text transcriptions into the Partitur-Editor

A. Preparing the file for import

There are two ways to import transcriptions of spoken language created with a text editor or a word processor (e.g. Word, OpenOffice etc.) into the Partitur-Editor:

- You can use the plain text (.txt) as input format. The plain text format will be useful if you have not been using any formatting information (e.g. instances of **bold** or *italic* text).
- Respectively, you can also use the "Simple EXMARaLDA" format. It is a format for plain text files. It can also handle transcriptions with some basic annotations, non-verbal behavior and overlapping speech.
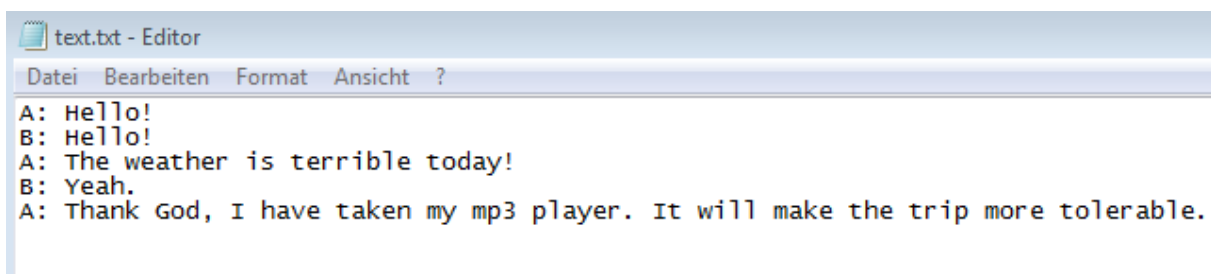
Please note: Fully automatic conversion of your transcription format into the Simple EXMARaLDA format is only possible if you have used layout and/or mark-up in a **consistent** way, with all different kinds of information encoded differently and in a way that can be recognized without human interpretation. Revisit your transcription key to see if any ambiguous annotations or mark-up need to be adjusted manually.

The Simple EXMARaLDA format can only handle the transcription itself – if you need to remove some preamble with e.g. metadata about the communication and the speakers, remember to save a copy of the transcription with the respective metadata.

B. Importing the file into the Partitur-Editor

Let's consider the following .txt file:

1. Click on **File > Import**

   a. In case you've decided to use the plain text format, choose the **Plain text file (*.txt)** from the dropdown menu (1). If necessary, change the options for character encoding (dropdown menu **Char encoding**) from **System-Default** to your preferred ones (2). However, if you don't know the character encoding, first try the default. Confirm your selection (3)
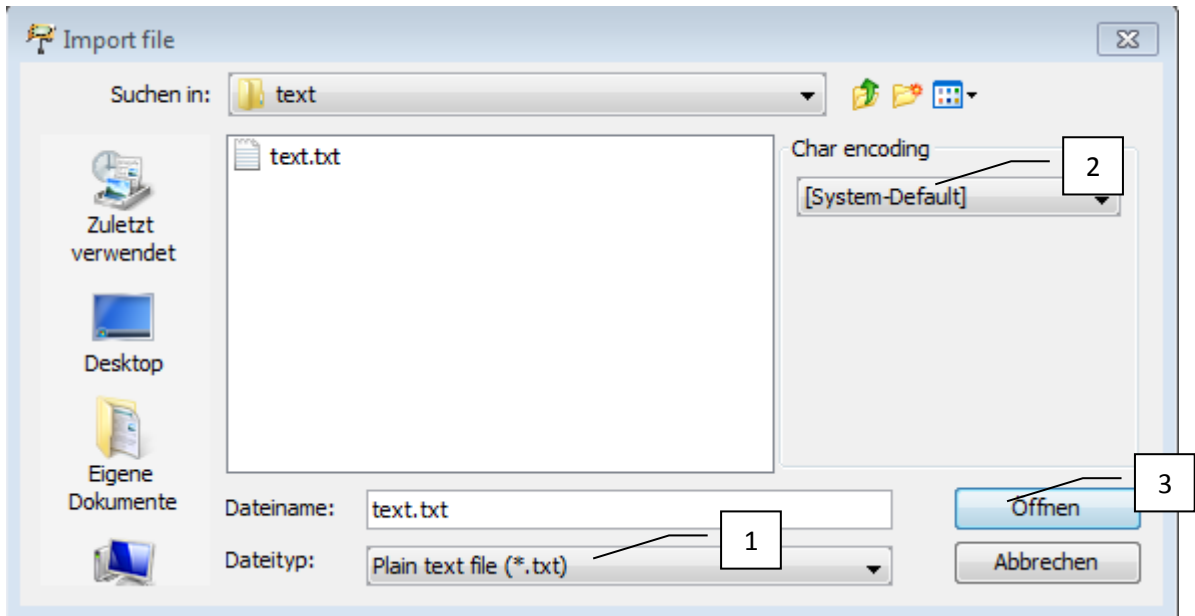
Then, you will be asked to select your preferred text splitter:
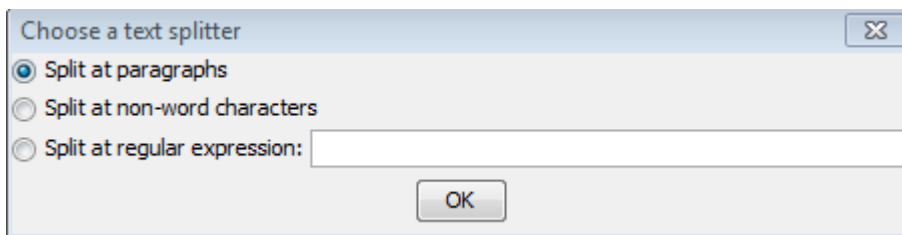
In more detail:

Split at paragraphs   All carriage returns will automatically generate event boundaries

Split at non-word characters   The text will be divided into word-alike chunks (non-alphabetic characters within words may be problematic), e.g.:


Split at regular expression   Event boundaries will be inserted at the exact points, at which the regular expression(s) you've typed matches your query. An example for a query "[\.\?]"

b. In case you will be using the Simple EXMARaLDA format, choose **Simple EXMARaLDA text file (*.txt)** from the dropdown menu (1). As already pointed out in section 1a. you should pay attention to **Char encoding** (2). Then click on **Open** (3):
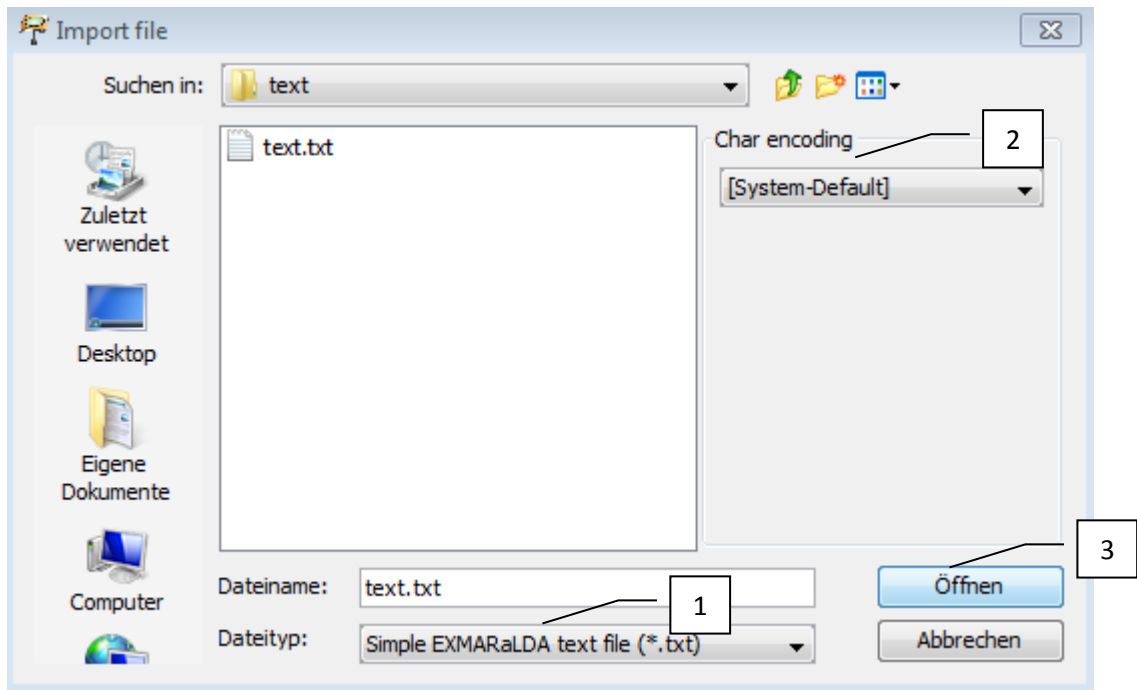
This time, the event boundaries will be inserted depending on the formatting conventions which were applied. In the example below using speaker abbreviations with a following colon (A:) is the simplest way of indicating a dialogue (note that it is still ambiguous whether there are overlaps or not):

A: Hello!
B: Hello!
A: The weather is terrible today!
B: Yeah!
A: Thank God, I have taken my mp3 player. It will make the trip more tolerable.

The same text, if imported to the Partitur-Editor as Simple EXMARaLDA format will look as follows:

The dialogue structure was converted into EXMARaLDA score.

2. Now you can save your file as **EXMARaLDA Basic Transcription (XML, EXB)** (1) **File > Save as...** (2). If you do not wish to save the formats (i.e. if you have edited the transcription), simply deselect **Save formats** (3):