Quickstart – Creating a corpus from a set of transcriptions

A. Preparation

The files of the EXMARaLDA Corpus on your hard drive should be organized as follows:

1. Media files should be located in the same folder or in a folder subordinate to the transcription they belong to.
2. Transcriptions that belong together (thematically or organizationally) should be held in the same directories.
3. The Coma file to be generated should be a level abovethese directories. In case of using the assistant "Create corpus from transcriptions...", this last point can be ignored as the file will automatically be located in the correct directory (s. B.).
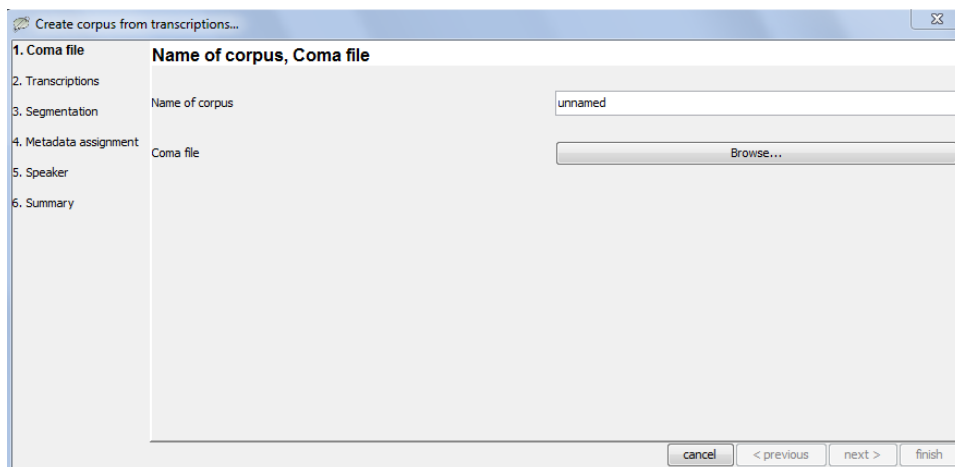
The files in the Partitur-Editor should be standardized as follows:

1. Abbreviations in the speakertable should be identical for each speaker.
2. Attribute labels in the speakertable for the same attributes should be identical.
3. Attribute labels in the metadata for the same kind of attribute should be identical.
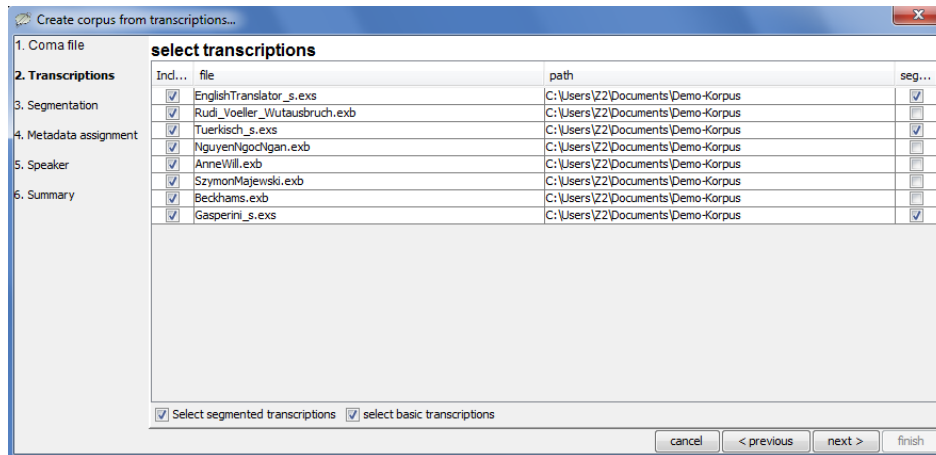
B. Using the assistant

This assistant is used to create a Coma file from a set of EXMARaLDA-transcriptions that already have metadata stored in their transcriptionheads. It will lead you through the necessary steps:

1. Open **COMA** and click on **File > Create a corpus from transcriptions...**. Enter the name for your corpus into the text field **Name of corpus**. Then click on the **Browse...**-button and navigate to the directory above all directories that contain transcriptions to be added to the corpus. The filename will automatically be created with the name you chose for thecorpus – if you want to change it, you can do it nowyou can now change it . Continue by clicking **next >**.
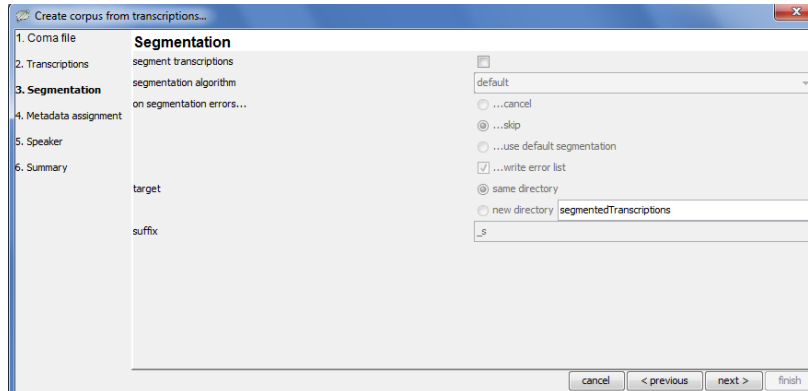
2. In this step, you can select the transcriptions to be included into the corpus. You can (de-) select single files with the checkboxes in the first column. The checkboxes in the last column are not clickable – they signal whether the transcription is a basic (unchecked) or segmented (checked) transcription. You can choose to include only segmented transcriptions by clicking on the "deselect all" and then on the "select basic transcriptions"-button.

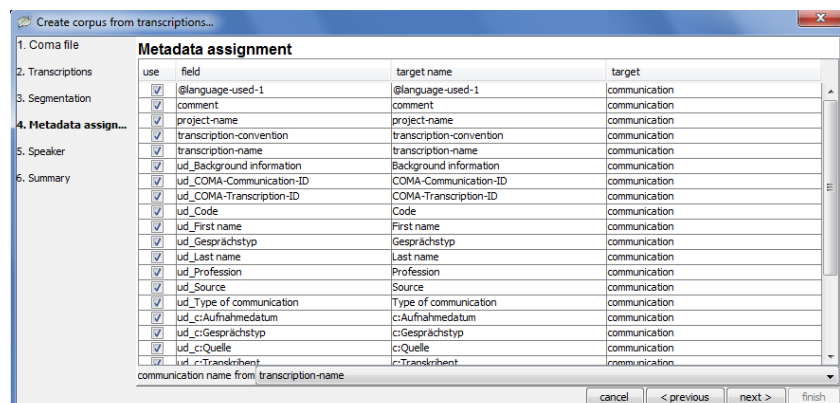
When you are done with your selection, click **next >**.

3. In this step you can decide to either create a segmented orbasic transcription while creating the corpus if you have not already done so. Further information about basic- and segmented transcriptions can be found in "Under-standing the basics of EXMARaLDA".
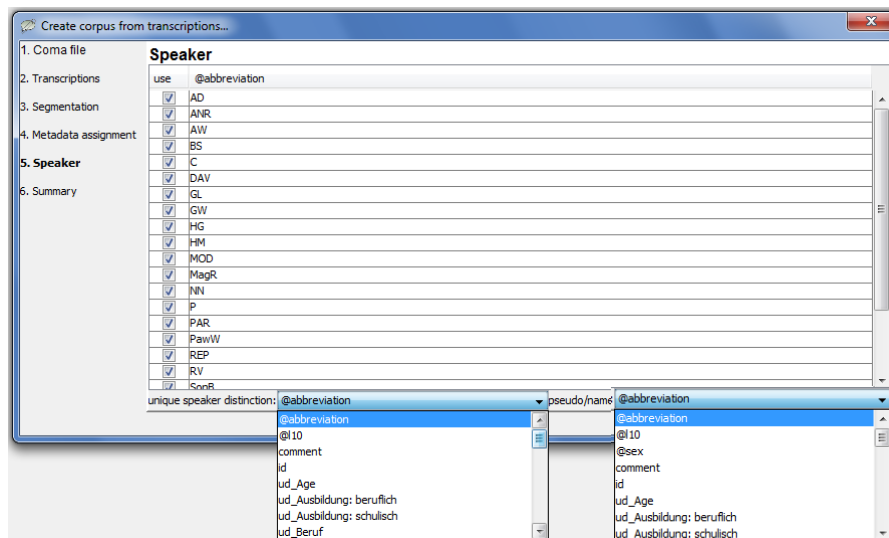

When you are done with your segmen-tation settings, click **next >**.

4. With **Metadata assignment** you decide, how the metadata found in the transcriptions (all fields found are shown in the second column of the table) will be used in the Coma-metadata-file.

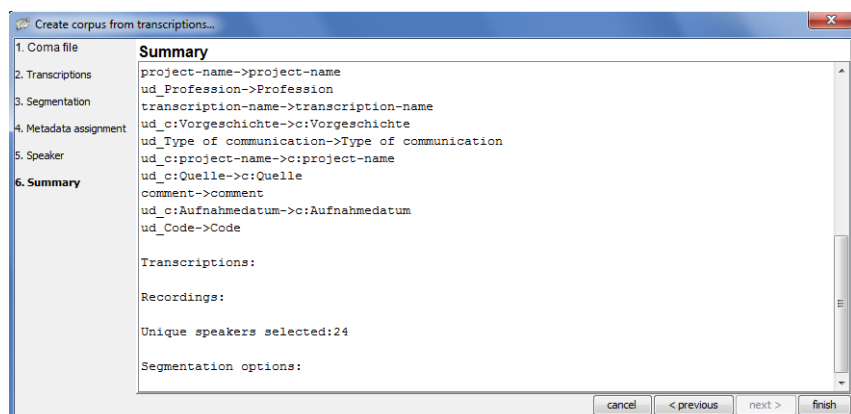If you are done editing the metadata assignments, click **next >**.

5. On the „**Speaker**-screen" you get a list of all speakers found in the speakertables of the found transcriptions. Now you can (de-)select which of the speakers, and via which meta-data field (use the drop down menu), shall be included in the COMA metadata.

When you are done with your speaker selection, click **next >**.

6. The next screen summarizes your choices on one screen. If you realize errors in the setup, just click on the "**< previous**"-button to correct the corresponding setting. When you are comfortable with your settings, click the **finish**-button.

If all goes well, the Coma-file is written and opened as an active document inside Coma right away.