

# How to use an EXMARaLDA corpus

This document explains how to use corpora from www.exmaralda.org. Examples are taken from the EXMARaLDA demo corpus. Other corpora work in a similar way but may differ, for example, with respect to the availability of audio data or the types of export formats.

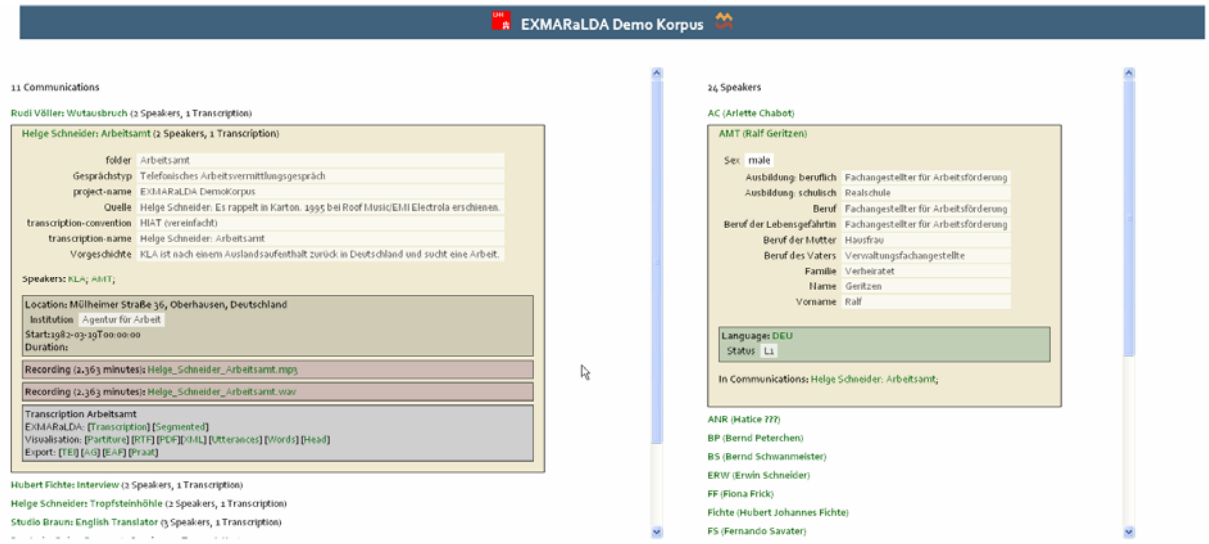
**Contents**

## *1. Online use*

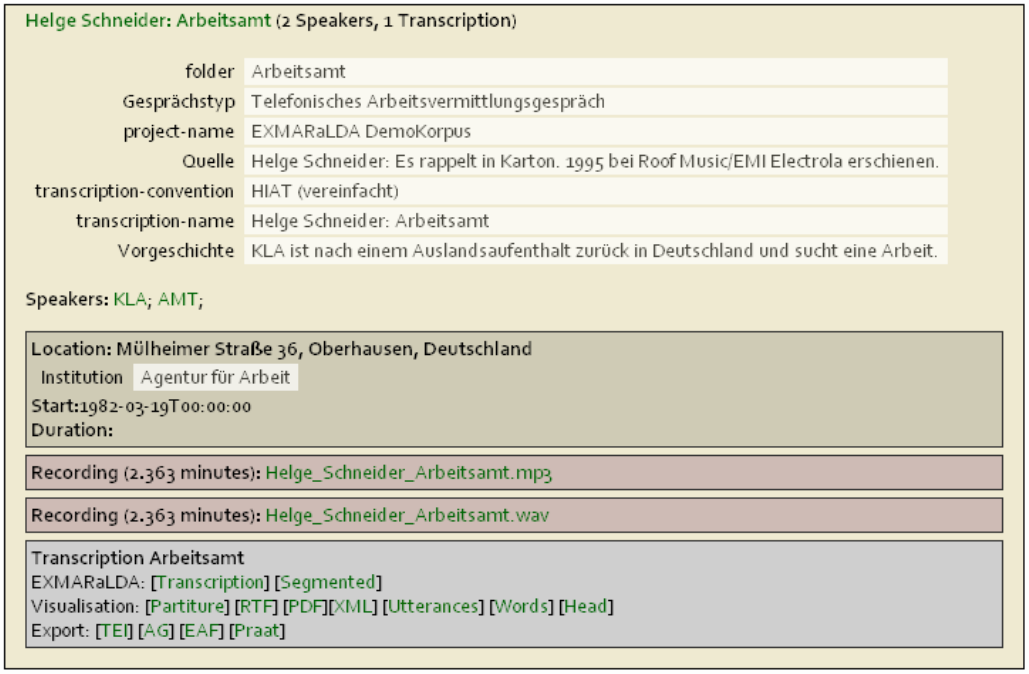
You can use a corpus online (i.e. using your internet browser and without downloading data) to browse meta data, transcriptions and recordings.

### Corpus overview

You usually start with a corpus overview:




The corpus overview consists of a list of communications (left side) and a list of speakers (right side). If you click on an item in one of these lists, information about the item will be displayed.




For communications, the top part gives you a list of meta data items. This is followed by a list of speakers participating in the communication. Click on any of the speakers to display the

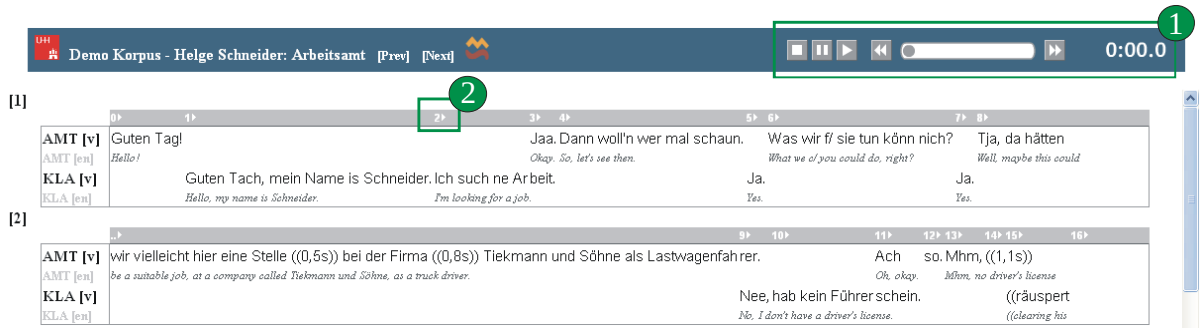
corresponding information in the speaker list. The lower part links to all the documents (recordings, transcriptions, visualisations and export formats) which belong to this communication. More specifically:

- The section **EXMARaLDA** links to an EXMARaLDA **Basic-Transcription**, which you can open and edit in the EXMARaLDA Partitur-Editor and an EXMARaLDA **Segmented-Transcription.**
- The section **Visualisation** links to **musical score** visualisations in four different formats (HTML, RTF, PDF and XML), to an **utterance list** (HTML) to a **word list** (HTML) and to a separate visualisation of the transcription **head** (HTML).
- The section **Export** links to several export formats. **TEI** is an XML file corresponding to the guidelines of the Text Encoding Initiative. **AG** is an annotation graph file which can be used for data exchange with various annotation tools. **EAF** is an ELAN Annotation File which can be opened and edited with the ELAN tool from the MPI in Nijmegen. **Praat** is a TextGrid which can be opened and edited with the Praat software.

Click on any of these files to display them in your browser. To download them you may have to right-click and choose "Download" from the context menu.

## HTML visualizations

If you display the HTML version of a musical score visualisation (and if the corpus makes the audio recordings available), you'll be given a transcription that is linked to a flash audio player (1):

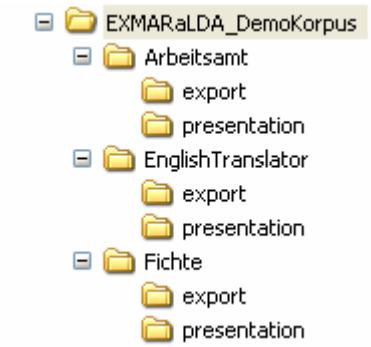



In the musical score, if you click on any one of the little arrows in the top rows (2), the player will start playback at the corresponding time in the audio recording. Clicking on any number in the top rows of a musical score will take you to the corresponding place in an utterance list:

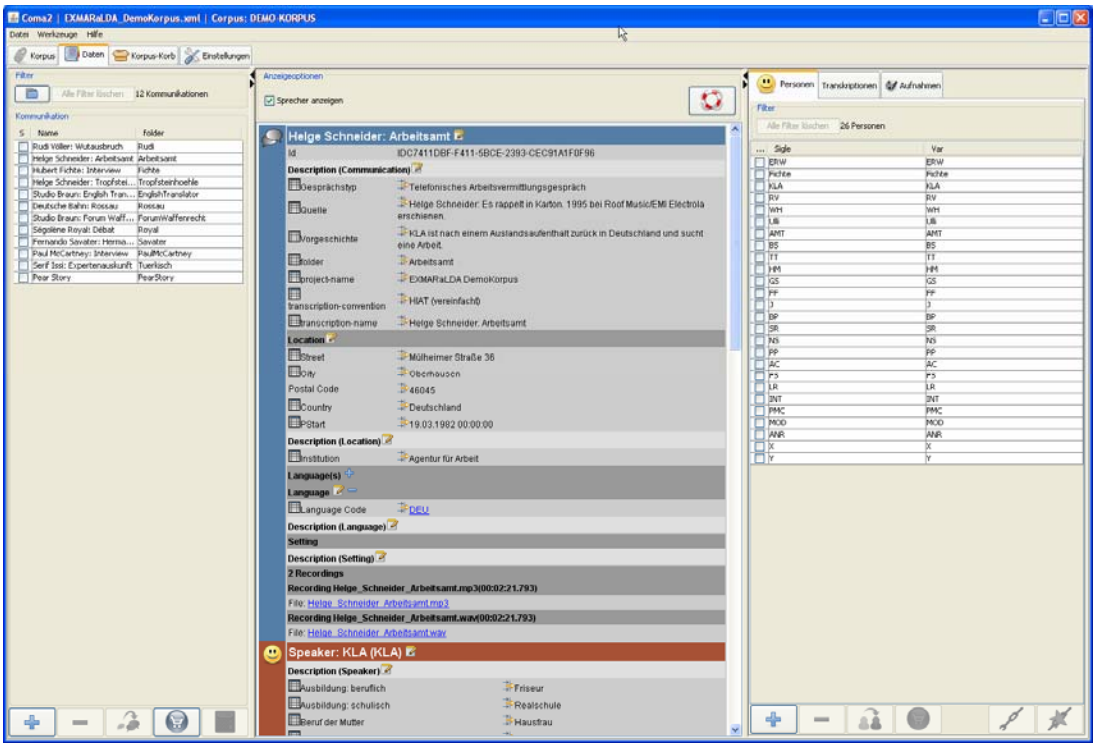
Here too, clicking on an arrow beside an utterance (3) will start the Flash Audio Player. Clicking on a number in square brackets will get you back to the corresponding part of the musical score visualization.

## *2. Offline use*

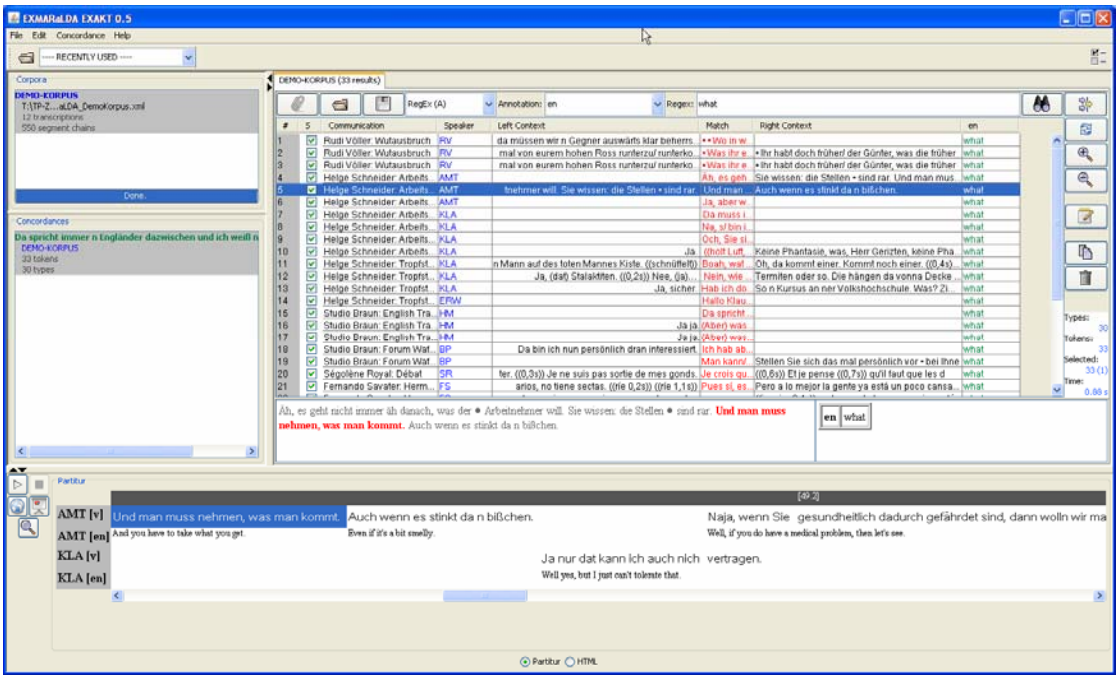
You can also download an entire corpus for offline use. This is especially useful if you want to edit data yourself or if you want to do corpus queries. To download a corpus, click on the link to the ZIP archive and unpack this archive on your hard disk. This should result in a directory structure like the following:




In the top level directory, there should be an XML corpus file (for the demo corpus, this file has the name EXMARaLDA_DemoKorpus.xml. You can open this file with the EXMARaLDA corpus manager to view, edit or query meta-data...




... or you can open the file with EXMARaLDA's query tool EXAKT to do corpus queries:




For further information on how to query an EXMARaLDA corpus, please consult the CoMa and EXAKT documentation. Of course, you can also view and edit individual transcription files with the EXMARaLDA Partitur-Editor.