



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 in a Browser	2
Corpus overview	2
HTML visualizations.....	3
2. Offline use	4
3. Use as remote corpus in EXAKT	6

1. Online use in a Browser

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 screenshot shows the EXMARaLDA Demo Korpus interface. On the left, there is a list of communications under the heading "11 Communications". The first item is "Helge Schneider: Arbeitsamt (2 Speakers, 1 Transcription)". Below this, a detailed view of the communication is shown, including fields for folder, Gesprächstyp, project-name, Quelle, transcription-convention, transcription-name, Vorgeschichte, Speakers, Location, Institution, Start, Duration, Recording, and Transcription. On the right, there is a list of speakers under the heading "24 Speakers". The first speaker is "AC (Arlette Chabot)", and below it, a detailed view of the speaker's information is shown, including fields for Sex, Ausbildung, Beruf, Beruf der Lebensgefährtin, Beruf der Mutter, Beruf des Vaters, Familie, Name, and Vorname. The interface also shows a list of other speakers at the bottom of the right panel.

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.

The detailed view of the communication "Helge Schneider: Arbeitsamt (2 Speakers, 1 Transcription)" is shown. It includes the following information:

- folder: Arbeitsamt
- Gesprächstyp: Telefonisches Arbeitsvermittlungsgespräch
- project-name: EXMARaLDA DemoKorpus
- Quelle: Helge Schneider: Es rappelt in Karton. 1995 bei Roof Music/EMI Electrola erschienen.
- transcription-convention: HIAT (vereinfacht)
- transcription-name: Helge Schneider: Arbeitsamt
- Vorgeschichte: KLA ist nach einem Auslandsaufenthalt zurück in Deutschland und sucht eine Arbeit.

Speakers: KLA; AMT;

Location: Mülheimer Straße 36, Oberhausen, Deutschland
Institution: Agentur für Arbeit
Start: 1982-03-19T00:00:00
Duration:

Recording (2.363 minutes): Helge_Schneider_Arbeitsamt.mp3
Recording (2.363 minutes): Helge_Schneider_Arbeitsamt.wav

Transcription: Arbeitsamt
EXMARaLDA: [Transcription] [Segmented]
Visualisation: [Partiture] [RTF] [PDF][XML] [Utterances] [Words]
Export: [TEI] [AG] [EAF] [Praat] [Chat] [FOLKER]

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** which is the file format used for querying in EXAKT.
- 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 (in older corpora) 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. **CHAT** is the file format of the CLAN editor of CHILDES. **FOLKER**, finally, is a file which can be opened with the FOLKER transcription editor of the IDS in Mannheim.

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):

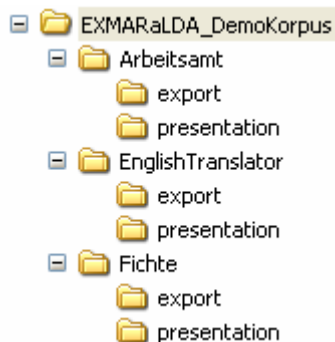
In the musical score, if you click on any one of the little arrows in the top rows (2), the player will start playback at the corresponding time in the audio recording. Clicking on any number in the top rows of a musical score will take you to the corresponding place in an utterance list:

Utterance	Speaker	Text	English Translation
1	AMT	Guten Tag!	Hello!
2	KLA	Guten Tach, mein Name is Schneider.	Hello, my name is Schneider.
3		Ich such ne Arbeit.	I'm looking for a job.
4	AMT	Jaa.	Okay.
5		Dann woll'n wer mal schaun.	So, let's see then.
6	KLA	Ja.	Yes.
7	AMT	Was wir f' sie tun könn nich?	What we c/ you could do, right?
8	KLA	Ja.	Yes.
9	AMT	Tja, da hätten wir vielleicht hier eine Stelle ((0,5s)) bei der Firma ((0,8s)) Tiekmann und Söhne als Lastwagenfahrer.	Well, maybe this could be a suitable job, at a company called Tiekmann und Söhne, as a truck driver.
10	KLA	Nee, hab kein Führerschein.	No, I don't have a driver's license.
11	AMT	Ach so.	Oh, okay.

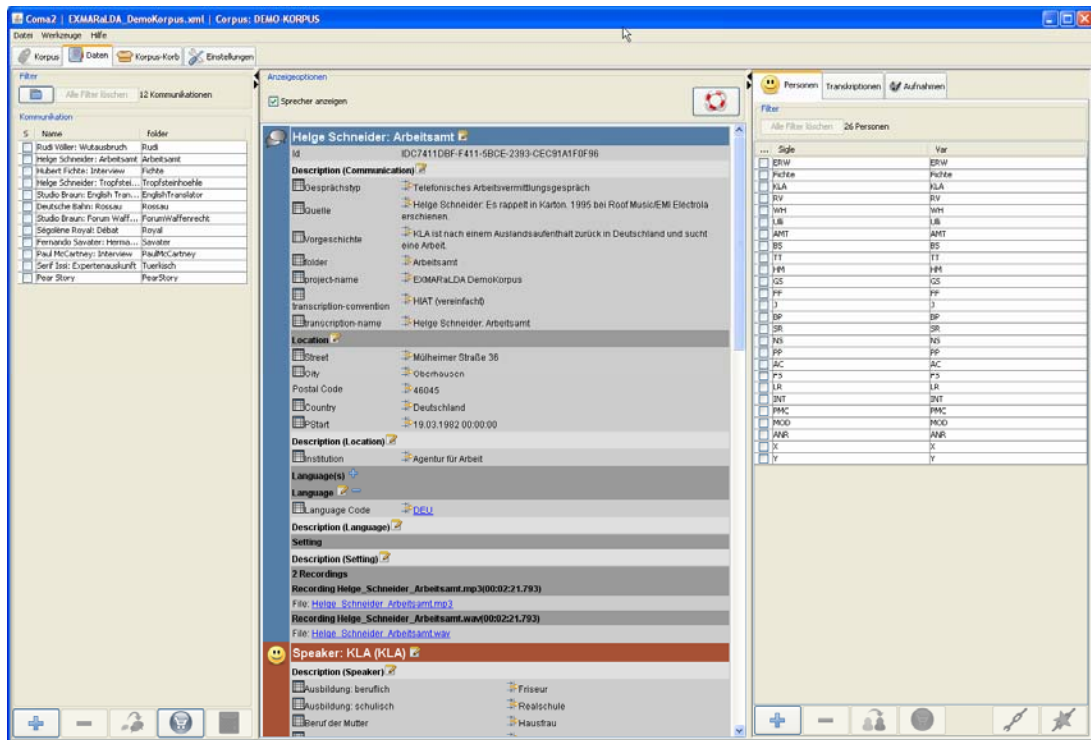
Here too, clicking on an arrow beside an utterance (3) will start the Flash Audio Player. Clicking on a number in square brackets will get you back to the corresponding part of the musical score visualization.

2. Offline use

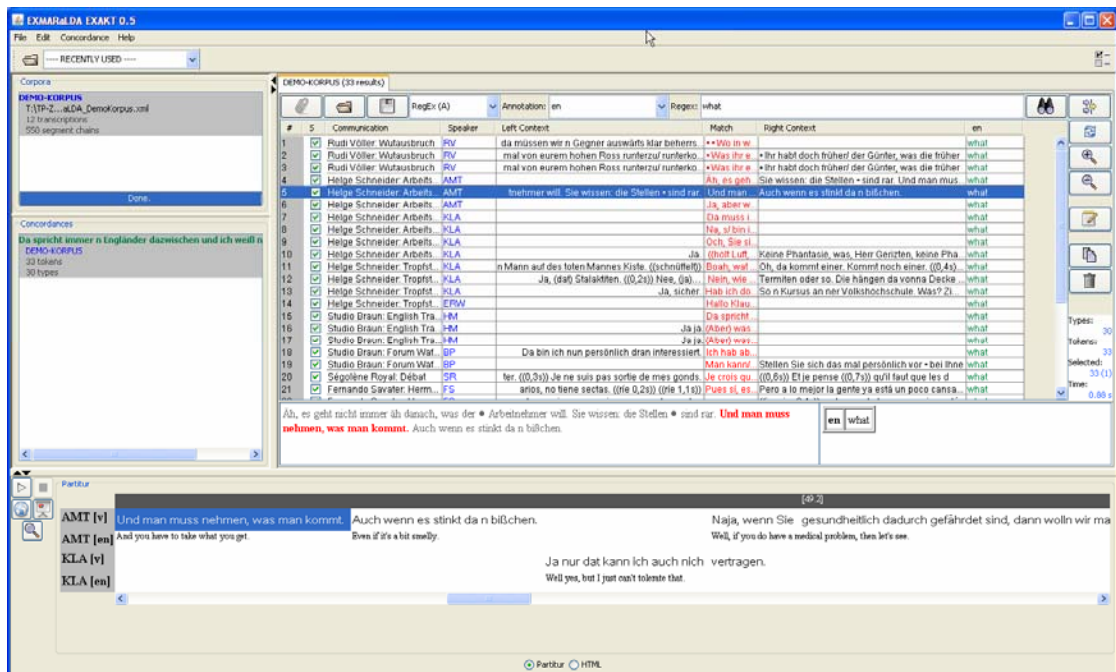
You can also download an entire corpus for offline use. This is especially useful if you want to edit data yourself or if you want to do corpus queries. To download a corpus, click on the link to the ZIP archive and unpack this archive on your hard disk. This should result in a directory structure like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



In the top level directory, there should be an corpus file with the suffix ".coma" (for the demo corpus, this file has the name EXMARaLDA_DemoKorpus.coma – older corpora use the suffix ".xml" instead of ".coma"). You can open this file with the EXMARaLDA corpus manager to view, edit or query meta-data...



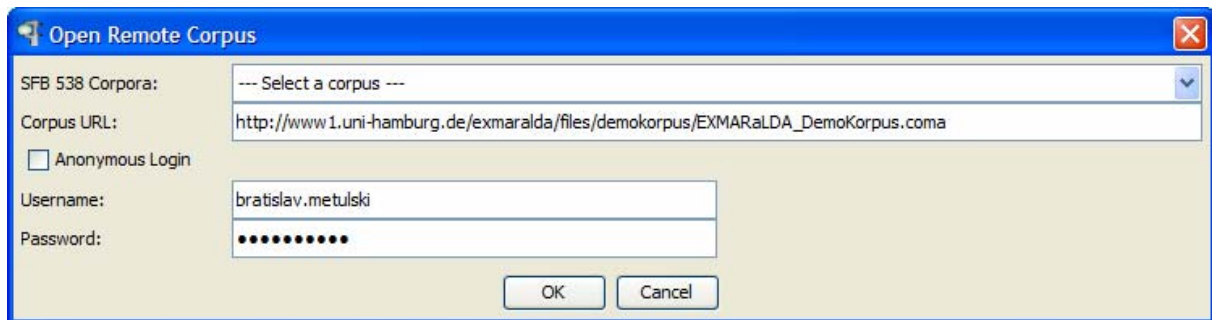
... or you can open the file with EXMARaLDA's query tool EXAKT to carry out corpus queries:



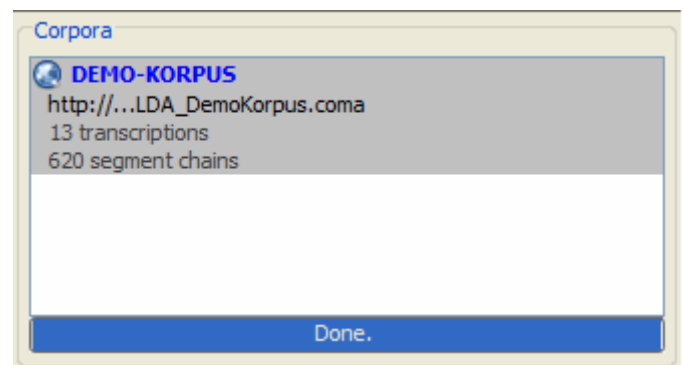
For further information on how to query an EXMARaLDA corpus, please consult the CoMa and EXAKT documentation. Of course, you can also view and edit individual transcription files (you'll find those inside the subfolders, they have the suffix ".exb" or – in older corpora ".xml") with the EXMARaLDA Partitur-Editor.

3. Use as remote corpus in EXAKT

Since version 0.7, EXAKT offers a possibility to do queries on a remote corpus, i.e. on a corpus that is not located on your own computer, but on a remote server. Note that you normally need a broadband connection for this feature to work satisfactorily (other connections are too slow). To open such a remote corpus, start EXAKT and choose **File > Open remote corpus...**



In the following dialog, you have to enter the URL address of the corpus' COMA file in the field **Corpus URL**. For corpora from the SFB 'Multilingualism', you can simply choose an entry from the field **SFB 538 Corpora** and the correct URL will be entered for you. If the corpus is password protected, you have to enter your username and password in the fields **Username** and **Password**, respectively. If the corpus is not password-protected (as is the case with the EXMARaLDA demo corpus), tick the box **Anonymous login** to disable user authentication. Clicking on OK will open the remote corpus and display it in the corpus list in the left upper corner of the EXAKT window.



You can then work with this corpus in the usual way, i.e. as if it was on your local computer. Note, however, that media playback for remote corpora is currently only supported on Windows. For Macintosh and Linux computers, media playback only works for local corpora.